SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(88th Edition)

APRIL 2020

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **Eighty Eight** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **April, 2020**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. 4-Bit Arithmetic Logic Unit That Uses Adjustable Threshold Pseudo-Dynamic Logic, ***7.Aman Agarwal***, ***8.Avijeet Kumar Trivedi***, ***8.Chirag Parashar*** and ***3.Neeta Pandey***, Electronics, DTU

2. A 1.1 lW biopotential amplifier based on bulk-driven quasi-floating gate technique with extremely low-value of offset voltage, Preeti Sharma, Kulbhushan Sharma, H. S. Jatana, ***6.1.Jaya Madan***, ***6.1.Rahul Pandey*** and Rajnish Sharma, A- Physics, DTU

3. A comparative analysis of illumination estimation based Image Enhancement techniques, ***7.Kavinder Singh*** and ***3.Anil Singh Parihar***, Computer Science, DTU

4. A Comparative Study of Subject-Dependent and Subject- Independent Strategies for EEG-Based Emotion Recognition using LSTM Network, ***7.Debarshi Nath[1]***, ***7.Anubhav[1]***, ***8..Mrigank Singh[1]***, ***3.Divyashikha Sethia[1]***, ***7.Diksha Kalra[2]*** and ***3.S. Indu[2]***, [1]Computer and [2]Electronics, DTU

5. A Multi-task Approach to Open Domain Suggestion Mining Using Language Model for Text Over-Sampling, ***6.Maitree Leekha*** , ***8.Mononito Goswami*** and ***3.Minni Jain***, Computer Science, DTU

6. A Novel Fuzzy based Intelligent Demand Side Management for Automated Load Scheduling, ***8.Dhairya Karna***, ***8.Aditya Vikram***, ***8.Astitva Kumar*** and ***3.M. Rizwan***, Electrical, DTU

7. A route selection approach for variable data transmission in wireless sensor networks, ***7.1.Aarti Jain***, Manju Khari, Elena Verdu, Shigeru Omatsu and Rube´n Gonza´ lez Crespo, Electronics, DTU

8. A study on the stability behavior of an epidemic model with ratio-dependent incidence and saturated treatment, ***8.Abhishek Kumar***, Manoj Kumar and ***3.Nilam***, Applied Mathematics, DTU

9. An Efficient Approach to EEG-Based Emotion Recognition using LSTM Network, ***7.Anubhav***[1], ***7.Debarshi Nath***[1], ***8.Mrigank Singh***[1], ***3.Divyashikha Sethia***[1], ***7.Diksha Kalra***[2] and ***3.S. Indu***[2], [1]Computer and [2]Electronics, DTU

10. An experimental analysis of tensile, hardness and wear properties of aluminium metal matrix composite through stir casting process, ***7.Anubhav Coyal***, ***3.N. Yuvaraj***, ***3.Ravi Butola*** and ***7.Lakshay Tyagi***, Mechanical, Production and Automobile, DTU

11. Analysis of Congestion Control Mechanism for IOT, ***6.Aastha Maheshwari*** and ***3.Rajesh Kumar Yadav***, Computer Science, DTU

12. Aqueous-Phased Electrospun Bioactive Glass Mineralized Gelatin-Pectin Hybrid Composite Fiber Matrix For 7- Dehydrocholesterol Delivery, ***6.Nidhi Gupta***, ***6.Himansh Goel***, ***3.Deenan Santhiya***, Chandra Mohan Srivastava, Sarita Mishra and ***6.Pragya Rai***, Applied Chemistry and Polymer Technology, DTU

13. Biofuel Certification Performance: A Review & Analysis, Sanjay Mohite and ***3.Sagar Maji***, Mechanical, DTU

14. Characterization of Line-Cut Signed Graphs, Rashmi Jain, Mukti Acharya and ***3.Sangita Kansal***, Mathematics, DTU

15. Classification of Cuisines from Sequentially Structured Recipes, ***8.Tript Sharma,*** Utkarsh Upadhyay and Ganesh Bagler, Mechanical, DTU

16. DESIGN AND FABRICATION OF WING OF A HIGH-PAYLOAD AERIAL VEHICLE, ***8.PRANJAL SHIVA***, ***8.GURU DEEP SINGH***, ***8.RUPANJALI KUKAL*** AND ***3.SANJAY KUMAR***, Mechanical, DTU

17. Electrochemical biosensor for the epithelial cancer biomarker EpCAM based on reduced graphene oxide modified with nanostructured titanium dioxide, ***6.Owais Jalil***, ***3.Chandra Mouli Pandey*** and ***3.Devendra Kumar***, Chemistry, DTU

18. FPGA Simulation of Fingertip Digit Recognition Using CNN, ***8.MD Shahbaz Khan***, ***6.Niharika***, ***6.Priya Yadav***, ***6.Rishabh Verma*** and ***3.Dr Indu Sreedevi***, Electronics, DTU

19. Handling Imbalanced Data using Ensemble Learning in Software Defect Prediction, ***3.Ruchika Malhotra*** and ***8.1.Juhi Jain***, Computer Science, DTU

20. Hierarchical Clustering of World Cuisines, ***8.Tript Sharma[1]***, Utkarsh Upadhyay, ***8.Jushaan Kalra[2]***, Sakshi Arora, Saad Ahmad, Bhavay Aggarwal and Ganesh Bagler, [1]Mechanical and [2]Computer, DTU

21. Hybrid Model for Lung Nodule Segmentation based on Support Vector Machine and k-Nearest Neighbor, ***8.Srishti Sharma[1]***, Prasenjeet Fulzele and ***3.Indu Sreedevi[2]***, [1]Electrical and [2]Electronics, DTU

22. Investigation of electrical/analog performance and reliability of gate metal and source pocket engineered DG-TFET, Jaya Madan, Rahul Pandey, Rajnish Sharma and ***3.Rishu Chaujar***, Applied Physics, DTU

23. Language Detection of Text Document Image, ***3.Dr. N. Jayanthi***, ***8.Harsha***, ***8.Naman Jain*** and ***8.Ishpreet Singh Dhingra***, ECE, DTU

24. MTFCT: A task offloading approach for fog computing and cloud computing, ***3.Rajni Jindal***, Neetesh Kumar and ***6.Hitesh Nirwan***, CSE, DTU

25. NLP based Machine Learning Approaches for Text Summarization, ***3.Rahul, Surabhi Adhikari*** and ***6.Monika***, Computer Science, DTU

26. Noisy three-player dilemma game: Robustness of the quantum advantage, ***8.Pranav Kairon***, Kishore Thapliyal, R. Srikanth and Anirban Pathak, Physics, DTU

27. Office Spirituality and Job pleasure: Exploring the Mediating Effect of Office Citizenship Conduct, ***8.1.Uday Sai***, USME, DTU

28. Rainfall Induced Landslide Studies in Indian Himalayan Region: A Critical Review, Abhirup Dikshit, ***3.Raju Sarkar***, Biswajeet Pradhan, Samuele Segoni and Abdullah M. Alamri, Civil, DTU

29. Real-time visual tracking viamulti-cue based adaptive particle filter framework, ***8.Ashish Kumar***, Gurjit SinghWalia and ***3.Kapil Sharma***, DTU

30. Resistorless current mode precision rectifier using EXCCII, Rupam Das and ***Sajal K. Paul***, Electronics, DTU

31. Resonant controller of single-stage off-board EV charger in G2V and V2G modes, ***6.Aakash Kumar Seth*** and ***3.Mukhtiar Singh***, Electrical, DTU

32. RESPONSE ANALYSIS OF PLUS SHAPED TALL BUILDING WITH DIFFERENT BRACING SYSTEMS UNDER WIND LOAD, ***3.Dr. Ritu Raj***, ***8.Shubhangi Jha***, ***8.Shreyansh Singh*** and ***8.Siddhant Choudhary***, Civil, DTU

33. Selection of sustainable solutions for crop residue burning: an environmental issue in northwestern states of India, ***3.Pravin Kumar*** and Rajesh Kumar Singh, Mechanical, DTU

34. Solar Thermal Collector Integrated Organic Rankine Cycle Technology, ***6.Harwinder Singh*** and ***3.R.S. Mishra***, Mechanical, DTU

35. Speaker Recognition using SincNet and X-Vector Fusion, ***8.Mayank Tripathi***, ***8.Divyanshu Singh*** and ***3.Seba Susan***, IT, DTU

36. STRENGTH AND DURABILITY PERFORMANCE OF FLY ASH BASED GEOPOLYMER CONCRETE USING NANO SILICA, Sameer Vyas, Mohammad Sameer, ***3.Shilpa Pal*** and Neetu Singh, Civil, DTU

37. Study on Adequacy of Functional Characteristics of a Typical Urban Waste Water Treatment Plant, ***7.Tarun Shokeen***, ***3.SK Singh*** and ***3.Geeta Singh***, Environmental, DTU

38. Submicron full-color LED pixels for microdisplays and micro-LED main displays, Xianhe Liu, Yuanpeng Wu, ***8.1.Yakshita Malhotra***, Yi Sun, Yong-Ho Ra, Renjie Wang, Matthew Stevenson, Seth Coe-Sullivan SID Member and Zetian Mi1, Electronics, DTU

39. Synergism of Recurrent Neural Network and Fuzzy Logic for Short Term Energy Load Forecasting, ***8.Srishti Sharma***, ***6.Rahul Kumar Agrawal*** and ***3.Madan Mohan Tripathi***, Electrical, DTU

40. The effect of dust grains on the Weibel instability in presence of large amplitude electrostatic waves, Ajay Gahlot, ***3.Suresh C. Sharma*** and Jyotsna Sharma, Physics, DTU

41. Unpaved test sections reinforced with geotextile and geogrid, ***6.Meenakshi Singh***, ***3.Ashutosh Trivedi*** and Sanjay Kumar Shukla, Civil, DTU

42. Using Hybridized techniques for Prediction of Software Maintainability using Imbalanced data, ***3.Ruchika Malhotra*** and ***3.Kusum Lata***, Computer Science, DTU

43. VDTA Based Fractional Order Floating Inductor and its Applications, ***8.Megha Jain***, ***8.Reshma S Kumar***, ***6.Lakshmi Nair*** and ***3.Rajeshwari Pandey***, Electronics, DTU

44. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model, ***7.Pranav Adarsh***, ***7.Pratibha Rathi*** and ***3.Manoj Kumar***, Computer Science, DTU

1. *Chancellor*

2. *Pro Vice Chancellor*	***2.1.*** *Ex Pro Vice Chancellor*
3. *Faculty*	***3.1.*** *Ex Faculty*
4. *Teaching-cum-Research Fellow*	***4.1.*** *Asst. Librarian*
5. *Alumni*	
6. *Research Scholar*	***6.1.*** *Ex Research Scholar*
7. *PG Scholar*	***7.1.*** *Ex PG Scholar*
8. *Undergraduate Student*	***8.1.*** *Ex Undergraduate Student*

2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)

4-Bit Arithmetic Logic Unit That Uses Adjustable Threshold Pseudo-Dynamic Logic

Aman Agarwal
Department of Electronics and Communication Engineering
Delhi Technological University (formerly Delhi College of Engineering)
Delhi, India
Email: amanagarwal702@gmail.com

Avijeet Kumar Trivedi
Department of Electronics and Communication Engineering
Delhi Technological University (formerly Delhi College of Engineering)
Delhi, India
Email: avijeet.trivedi@gmail.com

Chirag Parashar
Department of Electronics and Communication Engineering
Delhi Technological University (formerly Delhi College of Engineering)
Delhi, India
Email: mechiragp1997@gmail.com

Neeta Pandey
Department of Electronics and Communication Engineering
Delhi Technological University (formerly Delhi College of Engineering)
Delhi, India
Email: neetapandey@dce.ac.in

***Abstract*— In this paper, in order to improve the performance and power dissipation of a chip, we present an Arithmetic and Logic Unit (ALU) using the "Power-performance tunable Pseudo-Dynamic topology". To achieve this, we have designed basic logic gates such as AND, OR and XOR. The general idea of the logic was inspired from the pseudo - dynamic logic [1]. The input and output of this logic are compatible with static gates and can replace them with pseudo - dynamic gates. The pulse generation in the intermediate nodes decreases the power dissipation. This logic uses skewing supply voltage to tune the threshold voltage which provides performance and power dissipation controllability. Although owing to the larger number of transistors used for the same functionality, pseudo-dynamic topology acquires a larger footprint as compared to static CMOS topology, but its delay and overall power dissipation is certainly better than Static CMOS logic.**

For carrying out our simulations and transient analysis, we used Cadence® Virtuoso® Schematic Editor and ADEL Simulator.

Keywords—Pseudo-Dynamic Logic, Pseudo-Static Logic, Cadence Virtuoso, Tunable threshold, Power-Delay Product.

I. INTRODUCTION

An ALU (Arithmetic and Logical Unit) is a vital part of almost every computing device be it microprocessors, computers, embedded designs etc. All these devices have great application in the field of consumer electronics. An ALU is a combinational digital circuit used to perform arithmetic and logic operations on binary numbers. It represents the fundamental building block of the processing unit. As all the above-mentioned devices execute various operations, an ALU becomes a device which is used very frequently in a system [6], and thus must be made as efficient as possible.

The quality of a circuit is judged on the basis of a few performance metrics such as speed (delay), power and area. A combined parameter known as power-delay product (PDP) is also used to gauge the quality of a design. PDP has the added benefit of taking both delay as well as power into account, with a lower PDP being a desirable criterion [3].

The majority of the circuits designed today use static CMOS logic because of its inherent advantages, such as low static power dissipation, high noise margins, and large fan-out capability. Though newer technologies are always being worked upon in hopes of obtaining a better design. General structure of static CMOS based logic is shown in Fig. 1.

Obtaining energy efficient design while maintaining computational performance is one of the main pursuits of contemporary processor design. While scaling of transistor dimensions has long been a source of performance improvements in circuits, a corresponding increase in leakage current for decreasing technology nodes is present. With each process shrinkage, the leakage power dissipation becomes a greater component of the total power dissipation. Hence leakage power should not be neglected, and designs which lower the static power dissipation while maintaining functionality and performance should be pursued. Apart from the circuit topology, another method used to reduce the static power dissipation is threshold voltage alteration. An increase in threshold voltage reduces the static power dissipation and decrease in threshold voltage results in increase in speed. A popular method used for changing the threshold voltage of a transistor is to use substrate-bias effect. Due to this effect, any difference between the substrate and source voltage leads to a change in the threshold voltage.

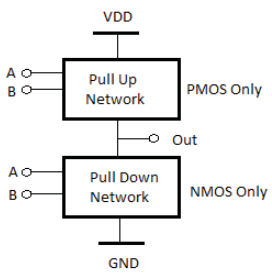


Fig. 1: Static CMOS logic general structure [3]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:45 UTC from IEEE Xplore. Restrictions apply.

II. RELATED WORK

In the majority of designs, there exists a trade-off between power, delay and area. Usually this tradeoff is decided at the time of manufacturing and cannot be altered later. In contrast to these designs, a logic family known as "pseudo-static logic" [2] was introduced. A top level block diagram of pseudo-static logic is shown in Fig. 2. Its path 1 propagates the rising input transition and path 2 propagates the falling input transitions. The weak path (path 3) at the bottom prevents the gates from becoming meta-stable and holds the output when the driving path is being reset. It allows the user to adjust the performance and power dissipation by adjusting ΔV (skew voltage). For positive ΔV, the overdrive voltage for the static OPL (Output Prediction Logic) block [5] decreases, resulting in increase in effective threshold voltage. This leads to lesser power dissipation and increased delay. Similarly, for negative ΔV, effective threshold voltage decreases resulting in increased power dissipation and smaller delay. This tunability provides a wider range of power-performance points for a design.

Another logic family, called adjustable threshold pseudo-dynamic logic family [1] was proposed in 2018. A top level block diagram of pseudo-dynamic logic is shown in Fig. 3. This logic structure was inspired by the pseudo-static logic and uses lesser number of transistors than it. The pseudo-static logic uses static OPL blocks that includes both pull up (PU) and pull down (PD) for each stage. The primary idea for decreasing the area in pseudo-dynamic logic is to replace the static OPL blocks used in pseudo-static logic with the dynamic one. In pseudo-dynamic logic the PD and PU of the CMOS are used in the first and the second stages respectively. Because of the dynamic OPL block used pseudo-dynamic logic family also incorporates tunability property in the same way as it is explained for the pseudo-static logic. Due to this tunability property, we can obtain a lower static power dissipation at the expense of reduced speed by increasing the threshold voltage, while increased speed can be obtained at the expense of increased static power dissipation by reducing the threshold voltage.

The pseudo-dynamic logic has pulses propagating in the internal nodes of the circuit while the external nodes operate with the same voltage levels as that of static CMOS circuits. The pulse generation in the intermediate nodes decreases the power dissipation, and since the external node signals are same in both static CMOS as well as pseudo-dynamic logic, they can be interchanged in a design without any modifications.

There are two different paths responsible for propagating the rising and falling input transitions. By combining the outputs of both the paths to a single output, the output node gets updated by either of the paths whenever an input transition occurs, while the other path's node signal remains unchanged. This leads to pseudo-dynamic circuits having the same activity factor as their static CMOS counterparts, consequently resulting in both logics having the same dynamic power dissipation.

In currently existing literature [1], buffer, AND, OR and XOR gates have been designed using the pseudo-dynamic logic. These circuits offer the same functionality as pseudo-static circuits while reducing the number of transistors used for same fan-in and thus having a smaller area. The pseudo-dynamic circuits also exhibit a lower power-delay product than their pseudo static-counterparts, effectively behaving as a better alternative to them.

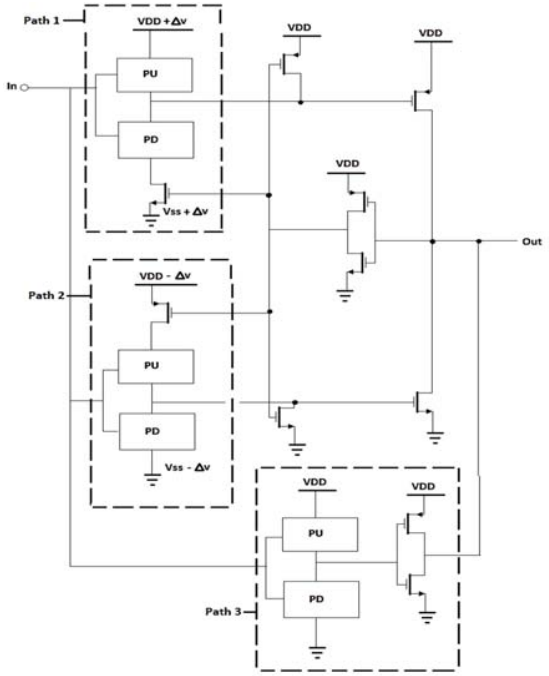


Fig. 2: Pseudo static logic general structure [2]

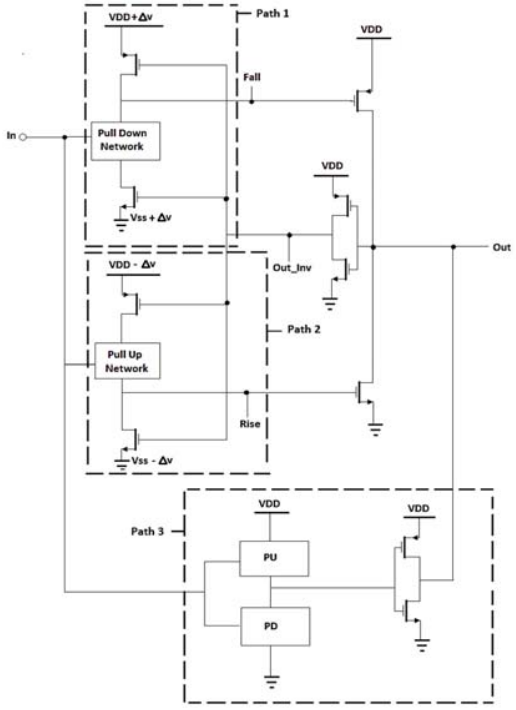


Fig. 3: Pseudo dynamic logic general structure [1]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:45 UTC from IEEE Xplore. Restrictions apply.

III. PROPOSED WORK

An ALU structure comprising of arithmetic and logic functions was designed using the pseudo-dynamic logic in an attempt to reduce the power dissipation, obtain better performance and also introduce tunability in the design. For this, initially the logic gates such as AND, OR, XOR were designed using the pseudo-dynamic logic (Fig. 3). These logic gates were then used to design the basic sub-blocks of an ALU such as multiplexers and full adders. The logic gates and sub-blocks were then connected to form the arithmetic block and logic block. A concatenation of these blocks' output through a 2:1 multiplexer formed the desired 4-bit ALU.

The schematic of designed 4-bit ALU is shown in Fig. 4. In order to compare the performance of ALU designed using pseudo-dynamic logic with that of static CMOS logic, the same ALU was also designed using the static CMOS logic by using the basic logic gates based on static CMOS logic. The goal was to obtain a design offering more power efficiency as well as better performance. To obtain the comparison, simulations were carried out to verify the functioning and measure static power dissipation and performance in Cadence® environment.

The truth table for ALU is given in table 1. It has four select lines, namely S3, S2, S1, S0. If S3 is low, arithmetic operations are carried out while if S3 is high, logic operations are carried out. By applying proper values to the select lines, arithmetic operations such as addition, addition with carry, subtraction, subtraction with borrow, increment, decrement, transfer and logic operations such as complement, AND, NAND, OR, NOR, XOR, XNOR can be performed. An additional operation to reset the output lines of ALU is also provided.

The gates designed using pseudo-dynamic logic are expected to consume lesser power and provide faster operation as compared to the static CMOS logic, owing to its internal pulse generation. The width of evaluation transistors is kept high to provide faster operation while the width of other transistors is kept small in order to reduce the static power dissipation.

A skew voltage, ΔV was also applied in the supply rails of the dynamic-OPL blocks of the logic gates designed using pseudo-dynamic logic (Fig. 3). It is expected that for more positive skew voltage applied in VDD and VSS, static power dissipation would reduce and propagation delay would increase due to a higher effective threshold voltage of the evaluation transistors. Similarly, for smaller skew voltages, static power dissipation would increase and propagation delay would reduce due to a reduction in the effective threshold voltage of the evaluation transistors.

By using the logic gates designed with pseudo-dynamic logic to build the ALU, we expect to obtain power and delay tunability by controlling the skew voltage in larger designs of pseudo-dynamic topology as well.

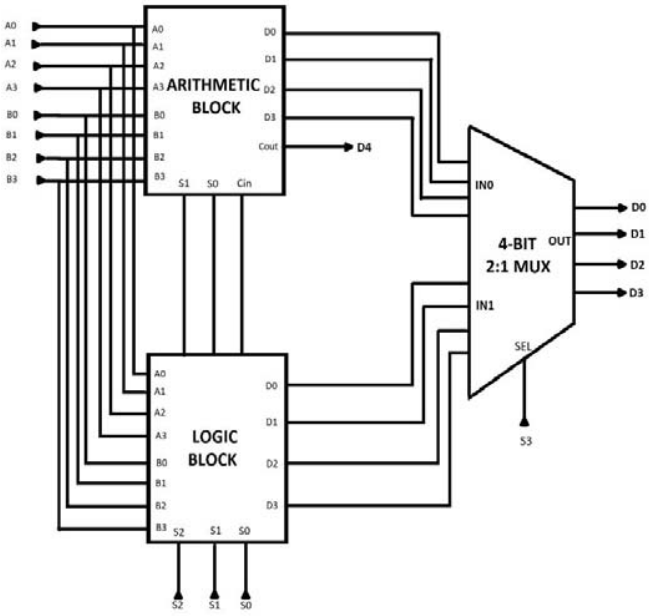


Fig. 4: ALU schematic

Table 1: ALU FUNCTIONALITY

S3	S2	S1	S0	F	FUNCTIONALITY
0	0	0	0	A+B	Add
0	0	0	1	A+B+1	Add with carry
0	0	1	0	A-B-1	Subtract with borrow
0	0	1	1	A-B	Subtract
0	1	0	0	A	Transfer A
0	1	0	1	A+1	Increment A
0	1	1	0	A-1	Decrement A
0	1	1	1	A	Transfer A
1	0	0	0	1	All 1
1	0	0	1	A'	Complement A
1	0	1	0	A ·B	AND
1	0	1	1	(A ·B)'	NAND
1	1	0	0	A ·B	OR
1	1	0	1	(A ·B)'	NOR
1	1	1	0	A ⊕ B	XOR
1	1	1	1	(A ⊕ B)'	XNOR

IV. RESULTS

Simulations were carried out using 180nm technology in Cadence® Virtuoso® for static CMOS and pseudo-dynamic logic. Results obtained are given in table 2-5. Table 2 shows the delay, power, and PDP for the basic logic gates designed using both static CMOS logic as well as pseudo-dynamic logic with zero skew (ΔV = 0). The results obtained verifies that the logic gates designed using pseudo-dynamic logic consumes lesser power and provides faster operation as compared to those designed using the static CMOS logic. Multiplying both propagation delay and power dissipation, we get power-delay product (PDP). PDP in pseudo-dynamic

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:45 UTC from IEEE Xplore. Restrictions apply.

logic gates come out to be substantially lesser than that of static CMOS logic gates.

The simulation results obtained for logic gates when skew is introduced are given in table 4. The skew voltages (ΔV) at which the outputs were taken are -0.1, 0, 0.1, 0.2 and 0.3. A graphical representation for the same is provided in Fig. 5. From the graphs obtained for OR gate, it can be observed that introduction of skew further reduces the static power dissipation by up to 30 percent, delay by up to 10 percent and PDP by up to 15 percent. Similar trends are also observed for AND and XOR gates.

Since, the arithmetic and logic unit is an interconnection of these basic logic gates, the PDP of the ALU designed using pseudo-dynamic logic is expected to be lower than that designed using the static CMOS logic. This is verified by the simulation results given in table 3. The results for pseudo-dynamic logic in table 3 are taken at zero skew voltage (ΔV). The simulation results of the pseudo-dynamic ALU with skew are given in table 5. The same skew voltages are applied for the ALU as well. Since in the ALU, the number of possible input combination is equal to 2^n, the delay and power will be different for different set of inputs. In our analysis, we have carried out simulations for the worst case. Worst case for delay means that we have taken inputs such that the carry is propagated to the most significant bit of the design. In case of power calculation, we have maximised the number of bits being toggled for the worst case value. The reported readings are the average of the delay and power dissipation for all the arithmetic and logic operations respectively. A graphical representation for the same is provided in Fig. 6. The same trend is observed for the ALU that was there in basic logic gates.

The metric in which pseudo-dynamic loses out to static CMOS is that of area. The pseudo-dynamic logic uses larger number of transistors for achieving the same functionality for same number of fan-in, which leads to this logic having a larger area as compared to static CMOS logic. The pseudo-dynamic logic has a greater transistor count for achieving the same functionality and fan-in than that of static CMOS. Transistor count for different logic gates are given in table 6. Fan-in for all the gates in table 6 is 2. It is clear from the table that the better PDP and tunabilty feature comes at the cost of more area on-chip per logic.

Table 2: POWER, DELAY AND PDP OF STATIC CMOS AND PSEUDO-DYNAMIC LOGIC GATES

Cell	Topology	Delay (ns)	Power (µW)	PDP (fJ)
OR	Static CMOS	0.109	0.286	0.0312
	Pseudo-Dynamic	0.090	0.267	0.0240
AND	Static CMOS	0.067	0.234	0.0156
	Pseudo-Dynamic	0.079	0.173	0.0115
XOR	Static CMOS	0.131	1.233	0.1610
	Pseudo-Dynamic	0.115	1.287	0.1484

Table 3: COMPARISION OF STATIC CMOS AND PSEUDO-DYNAMIC ALU

Unit	Topology	Average Delay (ns)	Average Power (µW)	PDP (fJ)
Arithmetic Functions	Static CMOS	1.1753	8.6390	10.1532
	Pseudo-Dynamic	0.8262	8.9820	7.4209
Logic Functions	Static CMOS	0.8390	3.3380	2.8006
	Pseudo-Dynamic	0.3924	3.5118	1.3779

Table 4: POWER AND SPEED FOR DIFFERENT SKEW OF LOGIC GATES

Gate	Skew	Average Delay (ns)	Average Power (µW)	PDP (fJ)
OR	-0.1	0.081	0.905	0.0733
	0.0	0.090	0.267	0.0240
	0.1	0.098	0.230	0.0225
	0.2	0.104	0.203	0.0211
	0.3	0.110	0.187	0.0205
AND	-0.1	0.062	0.198	0.0122
	0.0	0.079	0.173	0.0115
	0.1	0.083	0.161	0.0134
	0.2	0.094	0.151	0.0142
	0.3	0.099	0.141	0.0140
XOR	-0.1	0.111	1.386	0.1544
	0.0	0.115	1.287	0.1484
	0.1	0.161	1.197	0.1931
	0.2	0.169	1.078	0.1818
	0.3	0.176	0.936	0.1650

Table 5: POWER AND SPEED FOR DIFFERENT SKEW OF ALU

Unit	Skew	Average Delay (ns)	Average Power (µW)	PDP (fJ)
Arithmetic Functions	-0.1	0.757	9.786	7.4034
	0	0.826	8.982	7.4209
	0.1	0.900	7.477	6.7319
	0.2	0.977	7.076	6.9129
	0.3	1.058	6.945	7.3506
Logic Functions	-0.1	0.618	7.632	4.7202
	0	0.673	2.978	2.0048
	0.1	0.730	2.644	1.9301
	0.2	0.783	2.383	1.8659
	0.3	0.828	2.297	1.9008

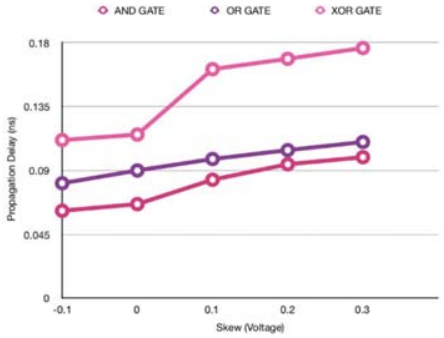
Table 6: NO. OF TRANSISTORS USED IN EACH TOPOLOGY

Cell	Topology	No. of Transistors
OR	Static CMOS	6
	Pseudo-dynamic	18
AND	Static CMOS	6
	Pseudo-dynamic	18
XOR	Static CMOS	8
	Pseudo-dynamic	30

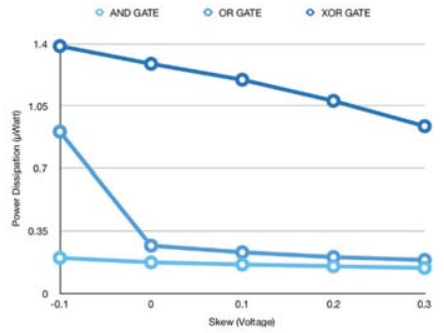
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:45 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION

The propagation delay obtained in all the circuits designed using pseudo-dynamic logic is smaller in comparison to static CMOS logic for the worst case analysis. The domain in which pseudo dynamic logic shines is the propagation delay metric, due to which the power-delay product of pseudo-dynamic logic is consistently lower than both static CMOS and pseudo-static logic. Also, the tunabilty offered by this topology can be exploited in mobile phones and laptops.

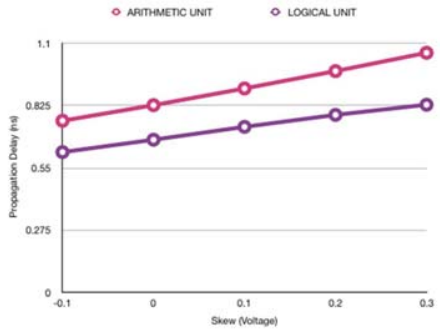


(a) Propagation delay of Logic gates wrt skew voltage

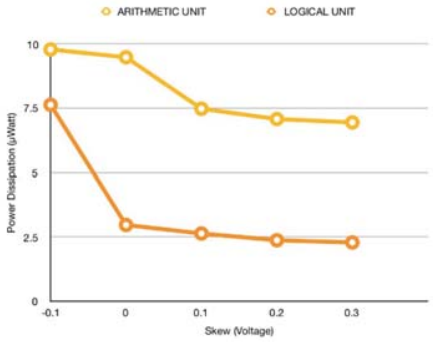


(b) Static power dissipation of Logic gates wrt skew voltage

Fig. 5: Variation of propagation delay and static power dissipation for different skew voltages in pseudo-dynamic based logic gates



(a) Propagation delay of ALU wrt skew voltage



(b) Static power dissipation of ALU wrt skew voltage

Fig. 6: Variation of propagation delay and static power dissipation for different skew voltages in the ALU

For applications which demand real time processing, the CPU can be tuned for high performance whereas applications that do not require fast processing, circuits can be tuned for power efficiency.

Although owing to the larger number of transistors used for the same logic functionality, pseudo-dynamic topology acquires a larger footprint as compared to static CMOS topology, but still lesser when compared to pseudo-static logic. All the resulting data points to the pseudo-dynamic logic being objectively better than static CMOS topology when area is not of concern.

As seen from the graphs, by varying the applied skew voltage, we can operate our circuits at a range of operating points by either finding an optimal balance between power dissipation and delay or by operating at points that favour either power dissipation or delay depending on the workload.

With every subsequent decrease in technology node size, the amount of total power made up by static power keeps on increasing. Due to which adopting this logic to design circuits with ever shrinking node sizes can be beneficial, allowing the user to have the advantages of both low power dissipation as well as faster processing.

VI. REFERENCES

[1] Gha□arishad H, Mohammadzadeh N, Ghaznavi□Ghoushchi M□B. A power□ performance tunable logic with adjustable threshold pseudo□ dynamic building blocks and CMOS compatibility. Int J Circ Theor Appl. 2018.

[2] Nezamfar B, Alon E, Horowitz M. Energy□performance tunable logic. IEEE Journal of, Solid□State Circuits. 2009;44(9):2554□2567.

[3] Rabaey JM, Chandrakasan AP, Nikolic B. Digital Integrated Circuits. 2 Englewood Cli□s: Prentice Hall; 2002.

[4] Kang S□M, Leblebici Y. CMOS Digital Integrated Circuits. New York: Tata McGraw□Hill Education; 2003.

[5] L. McMurchie, et al., "Output Prediction Logic: A High-Performance CMOS Design Technique," Proc. ICCD, pp. 247-254, Sept., 2000.

[6] Mano, Morris M. Computer System Architecture, 3e: Pearson Education Inc; 2007.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:45 UTC from IEEE Xplore. Restrictions apply.

Analog Integrated Circuits and Signal Processing (2020) 103:303–313
https://doi.org/10.1007/s10470-020-01623-y

A 1.1 μW biopotential amplifier based on bulk-driven quasi-floating gate technique with extremely low-value of offset voltage

Preeti Sharma[1] · Kulbhushan Sharma[1] · H. S. Jatana[2] · Jaya Madan[1] · Rahul Pandey[1] · Rajnish Sharma[1]

Received: 21 August 2019 / Revised: 30 January 2020 / Accepted: 4 March 2020 / Published online: 2 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
Biopotential amplifier (BPA) remains one of the most crucial blocks for the successful implementation of any of the biomedical systems. However, design of a BPA remains challenging owing to most of the topologies reported in literature displaying high values of noise, consuming high value of power and working in limited range of bandwidth. Thus, circuit topologies capable of providing an optimum and an acceptable combination of these parameters remains a topic of immense interest among researchers. We in this paper, report the results of a BPA designed using Bulk-Driven Quasi-Floating Gate (BDQFG) technique with a special focus on the effect of variation in the values of biasing resistor (R_{large}). Results obtained through mathematical modelling and analysis of the circuit have been verified by conducting simulations in Cadence Analog Design Environment using standard 0.18 μm technology. Circuit design has been optimized for least values of power consumption of the order of 1.1 μW, noise ($\approx$ 3.15 μV_{RMS}), mid-band gain of 39.9 dB (from 0.266 Hz to f_{-3dB} of 2.8 kHz), offset voltage of 455 μV and phase margin of 65.83°.

Keywords Biopotential amplifier · Bulk driven · Low power · Operational transconductance amplifier · Quasi-floating gate

1 Introduction

The next generation of biomedical systems demand for the suitable amplification of biomedical signals coming in from diagnosing equipments like Electroencephalogram (EEG), Electrocorticogram (ECoG), Electromyogram (EMG) and Electrocardiogram (ECG). All these signals possess small values of amplitude of the order of 1 μV to 5 mV with frequency values between 0.01 Hz to 7.2 kHz [1–4]. More precise analysis of biomedical signals helps in early diagnosis of chronic diseases like paralysis, arrhythmias, respiratory disorders, body fluid overload, epilepsy and parkinson etc. [5–7]. Very specifically, EEG, ECoG, EMG and ECG signals occupy the frequency band of 0.5–100 Hz, 0.5–200 Hz, 6–500 Hz and 0.01–250 Hz respectively [1, 3]. Faithful amplification of the same remains a major challenge owing to such small values of amplitude and frequency associated with these signals.

Other than this, most of the topologies reported in literature [5, 8, 9, 10, 11] pertaining to design of a BPA seem to suffer from major challenges like high values of total-input referred noise ($V_{in,RMS}$), power consumption and operable in limited range of bandwidth. Various closed loop amplifier architectures like Capacitively Coupled-Capacitive Feedback (CC-CF) [10], Directly Coupled-Active Feedback (DC-AF) [12] and Chopper Stabilized (CHS) [5] have been proposed by researchers to resolve these issues. All these reported architectures have their own set of strengths and limitations. While, CC-CF architecture

✉ Kulbhushan Sharma
kulbhushan.sharma@chitkara.edu.in

✉ Jaya Madan
jaya.madan@chitkara.edu.in

✉ Rahul Pandey
rahul.pandey@chitkara.edu.in

✉ Rajnish Sharma
rajnish.sharma@chitkarauniversity.edu.in

Preeti Sharma
preeti.s@chitkara.edu.in

H. S. Jatana
hsj@scl.gov.in

[1] VLSI Centre of Excellence, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

[2] Semi-conductor Laboratory (SCL), Mohali, Punjab, India

finds its strength in providing high values of gain and Common Mode Rejection Ratio (CMRR) [13], DC-AF and CHS architectures establish their advantage in terms of DC offset rejection [14, 15].

Based on the literature survey [12, 13, 14, 15, 16, 17, 18, 19, 20, 21], it can be established that the choice of Operational Transconductance Amplifier (OTA) topology and use of the type of pseudo-resistor therein majorly contributes to noise and power in the overall architecture of any BPA. Earlier reported research work has been focused on Current mirror [3, 16], Miller [17], Folded-Cascode [11], Telescopic [20], Differential self-bias [18], Floating-gate [21] and Complementary Metal Oxide Semiconductor (CMOS) inverter-based [22] OTA topologies to meet the requirements set for design of a close to ideal BPA. The choice of pseudo-resistor also helps in accurate stabilization of lower cut-off frequency (f_L) of a neural amplifier and saving on overall chip area [4].

Khateb et al. [23, 24] worked upon low-voltage and low-power analog circuits using Bulk-Driven (BD), Quasi-Floating Gate (QFG), Bulk-Driven Floating Gate (BDFG) and Bulk-Driven Quasi-Floating Gate (BDQFG) techniques to reduce power-consumption and noise. Among all these researched techniques, although BDQFG is found to occupy large chip area owing to usage of biasing resistor and coupling capacitor, but it still finds merit in reduced power consumption and noise. That is basically due to the combined advantages of bulk-driven and quasi-floating gate techniques in terms of higher values of the transconductance at low supply voltages. BDQFG technique has been used by Khateb et al. [24] and by Kumngern et al. [25] to design a rectifier and current conveyor respectively.

We in this paper report the design of a BPA comprising of an OTA and Psuedo-Resistor (PR) based on Bulk-Driven Quasi-Floating Gate (BDQFG) technique. A series of analysis based simulations have been carried out to understand the effect of variation of values of biased resistance in BDQFG transistors on the overall design of BPA. Optimized simulation results reveal the high value of the gain of the order of 39.9 dB, least amount of power consumption equal to 1.1 μW, total input-referred noise ($\approx$ 3.15 μV_{RMS}) and offset voltage with the value of 455 μV for the proposed technique.

Paper has been divided in four different Sections other than providing the list of references at the end. Section 2 providing all the details about the design and simulation methodology adopted for this research effort follows the present section of Introduction. Results in the form of simulation plots ably supported by the justification for the same are presented in "Results and Discussions", Sect. 3. Paper culminates in providing the conclusion of the research effort in Sect. 4.

2 Design and simulation methodology

It is necessary to follow a design methodology before the direct implementation of circuit on Virtuoso Analog Design Environment tool so as to meet the desired specifications for design of a BPA such as high value of gain and low values of input-referred noise voltage, power and offset voltage. Details of the same have been presented in this section of the paper. A detailed transconductance analysis and noise analysis of the used OTA topology have been carried out so as to study the effect of variation of bias resistor (R_{large}).

2.1 Low power biopotential amplifier

Schematic shown in Fig. 1 shows generic CC-CF architecture of a BPA typically highlighting the presence of an OTA block, PRs (R_F) and Capacitances (C_{in}, C_F and C_L). Similar basic configuration has been worked upon earlier by Harrison et al. [9] where they have used a series combination of transistors of same size to provide a certain value of resistance at the feedback loops.

Primarily C_{in} and C_F play a major role in controlling the mid-band gain (A_1) owing to the basic relationship mentioned in Eq. (1).

$$A_1 = \frac{C_{in}}{C_F} \tag{1}$$

Band-limiting load capacitance (C_L) helps in fine tuning the overall −3 dB bandwidth (f_{-3dB}) which is approximated by the Eq. (2).

$$f_{-3dB} = \frac{g_{m,OTA}}{2\pi A_1 C_L} \tag{2}$$

where, $g_{m,\ OTA}$ is the transconductance of the OTA. C_F and R_F connected in the feedback path of OTA are responsible for determining the f_L of the BPA as given in Eq. (3).

$$f_L = \frac{1}{2\pi R_F C_F} \tag{3}$$

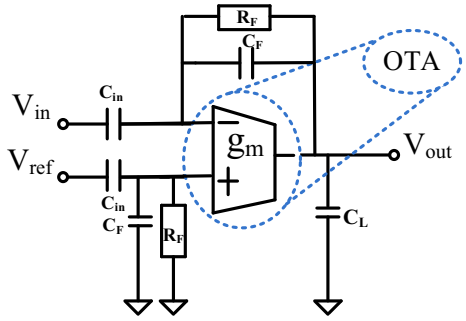


Fig. 1 Schematic of generic CC-CF biopotential amplifier architecture

Another set of R_F and C_F connected at the non-inverting terminal of OTA is responsible for basic filtering action and weed out the unnecessary DC components present in the input signals [9]. The details of design methodology of internal architecture of BPA like OTA and PR has been provided in the following Sections.

2.1.1 OTA topology

Elaborated circuit schematic used for realization of OTA mentioned in Fig. 1 has been shown in Fig. 2. Important specifications for carrying out the design of an OTA for a BPA application like total-input referred noise voltage (1–4 μV_{RMS}), power consumption ($< 10\ \mu W$), gain (~ 40 dB), stability (PM $> 60\ ^\circ$) and unity gain bandwidth product (~ 75 kHz) have been considered for finding out the required design parameters [26]. The aspect ratio to ascertain the size of the active devices (M_0-M_9) used in the BDQFG OTA and ensure their working in the right operating regions for satisfactory working of BDQFG OTA has been performed with the help of Eqs. (4–6) earlier used by Harrison et al. [9].

$$IC = \frac{I_D}{I_S} \tag{4}$$

IC is inversion coefficient; I_D is drain current and I_S is moderate-inversion characteristic current.

Values of I_S and I_D can be found by Eqs. (5) and (6) as follows:

$$I_S = \frac{2\mu_p C_{ox} V_T^2 W}{kL} \tag{5}$$

μ_p is hole mobility, C_{ox} is oxide capacitance, k (= 0.7) is a constant

$$I_D = \frac{W}{L} I_{D0} e^{\frac{(V_{GS}-|V_{TH}|)}{\eta V_T}} \tag{6}$$

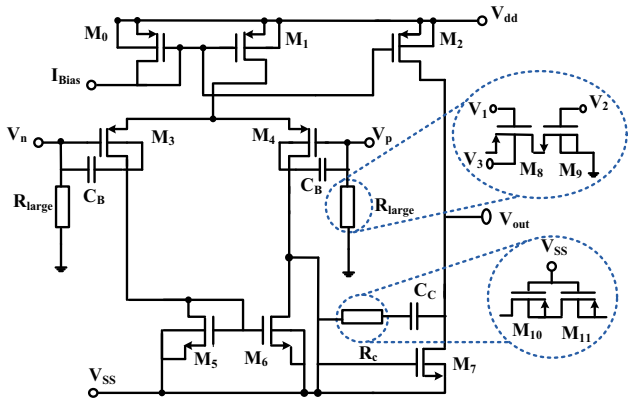


Fig. 2 Schematic of BDQFG OTA: the transistors M_3 and M_4 have been biased using resistor R_{large}

I_{D0} is process-dependent parameter; η is slope-factor; V_{TH} is the transistor's threshold voltage; V_T is the thermal voltage (26 mV); V_{SG} is the source-to-gate voltage.

The transconductance (g_m) of gate- driven transistors (M_0-M_9) used in the BDQFG OTA can be calculated by using Eq. (7):

$$g_m \approx \frac{kI_D}{V_T} \frac{2}{1+\sqrt{(1+4IC)}} \tag{7}$$

Important design parameters thus found like aspect ratios of active devices and their recommended regions of operation along with values of capacitors, power supply and pseudo-resistors adopted in Figs. 1 and 2 have been provided in Table 1.

A. Transconductance Analysis

The devices M_3 and M_4 used in the differential amplifier stage of OTA have been biased using BDQFG approach, which provides higher value of g_m along with low power consumption and noise in comparison to bulk-driven, floating-gate and quasi-floating gate techniques [27]. The input signals V_p and V_n are applied to the bulk and quasi-floating gate terminal of M_3 and M_4 respectively through coupling capacitor (C_B) connected between bulk and gate. Biasing to these transistors has been provided through MOS based complimentary pseudo-resistors, R_{large} = 9.1 TΩ. A series of simulations to optimize this value of R_{large} have been conducted as elaborated in Sect. 3. Details on the design of this resistor can be found in earlier reported work by Dehsorkh et al. [28]. This kind of circuit arrangement leads to enhanced values of transconductance given by Eq. (8) [27].

$$g_{m,BDQFG} = g_{mb} + g_{m,QFG} \tag{8}$$

In Eq. (8), $g_{mb} = (0.2\text{–}0.4)g_m$ and $g_{m,QFG} = (C_B/C_{Total})g_m$ indicate the transconductances of bulk-driven MOSFET and quasi-floating gate MOSFET respectively. Kumngern et al. [25] demonstrated that the R_{large} and total-internal capacitances, C_{Total} used in the BDQFG transistor act as the high-pass filter. Equation for cut-off frequency is found in (9) [25].

$$f_{cut,off} = \frac{1}{2\pi R_{large} C_{Total}} \tag{9}$$

Substituting the value of C_{Total} from Eq. (9) into (8), the modified expression for $g_{m,BDQFG}$ be found in (10):

$$g_{m,BDQFG} = g_{mb} + \left(2\pi R_{large} f_{cut,off} C_B\right) g_m \tag{10}$$

From (10), it can be clearly established that g_m of BDQFG topology is directly dependent on the value of R_{large}. Accordingly, the enhanced values of $g_{m,BDQFG}$ result in better values of unity gain bandwidth product (UGBW).

Table 1 Aspect ratios, operating region of devices and component values used in the biopotential amplifier circuit reported in this work

Devices	W/L (μm/μm)	IC	I_D (nA)	g_m (μS)	g_m/I_D (V^{-1})	Region of operation
M_0, M_1	1/1	3.13	498	6.04	12.50	Saturation
M_2	1/1	1.1	498.2	6.28	12.60	Saturation
M_3, M_4	600/8	0.03	241.3	6.41	26.58	Subthreshold
M_5, M_6	48/14.2	0.03	242	5.56	22.95	Subthreshold
M_7	1/1	3.44	498.2	9.313	18.7	Saturation
M_8, M_9	600/180	0.022	1.308 f	11.46 f	8.76	Subthreshold

V_{dd} = 500 mV, C_C = 3.5 pF, C_{in} = 30 pF, C_F = 300 fF, C_B = 5 pF, C_L = 100 fF, $R_F \approx 2$ TΩ, Vss = -500 mV, I_{Bias} = 500 nA, $R_C \approx 400$ GΩ, $R_{large} \approx 9.1$ TΩ, V_1 = -0.1 V, V_2 = -0.2 V, V_3 = 1.8 V, V_{ref} = 0 V

With an understanding about the possible approach to improve the characteristics of the differential amplifier used in Fig. 2, a small-signal model of the complete OTA has been drawn as shown in Fig. 3. The overall transconductance of the proposed OTA ($g_{m,OTA}$) is obtained by the Eqs. (11–13).

$$g_{m,OTA} = \frac{(g_{mb}+g_{m,QFG})(1-g_{mn0}R_C)R_1C_C\left(s-\frac{g_{mn0}}{C_C(1-g_{mn0})}\right)}{1+[R_1C_C+R_1(C_{gsn0}+C_C)]s+[R_1(R_CC_{gsn0}C_C)]s^2} \tag{11}$$

Here, R_1 is the parallel resistance combination of the output resistance of M_6 (r_{o1}) and M_3 (r_{o3}), C_{gsn0} is the gate-to-source capacitance of M_7, R_C is the compensation resistor comprising of two series connected tunable PMOS devices (M_{10}, M_{11}), and g_{mn0} is the transconductance of M_7. The Eq. (11) can be rearranged into (12) as follows:

$$g_{m,OTA} = \frac{(g_{mb}+g_{m,QFG})(1-g_{mn0}R_C)R_1C_C\left(s-\frac{g_{mn0}}{C_C(1-g_{mn0})}\right)}{s^2+\left[\frac{1}{R_1C_{gsn0}}+\frac{C_{gsn0}+C_C}{R_CC_CC_{gsn0}}\right]s+\left[\frac{1}{R_1R_CC_CC_{gsn0}}\right]} \tag{12}$$

At s = 0, the transconductance can be obtained by the Eq. (13):

$$g_{m,OTA} = (g_{mb}+g_{m,QFG})g_{mn0}R_1 \tag{13}$$

It is clearly inferred from (13) that the transconductance of the proposed OTA is enhanced by the factor of $g_{mb} + g_{m,QFG}$.

The next parameter of OTA is unity gain bandwidth product ($UGBW_{OTA}$) that can be estimated using Eq. (14).

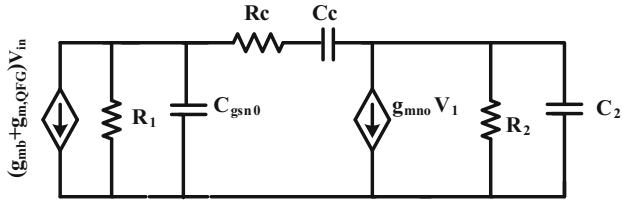


Fig. 3 Small-signal model of the reported OTA

$$UGBW_{OTA} = \frac{(g_{mb}+g_{m,QFG})}{C_C} \tag{14}$$

The enhancement in the transconductance ($g_{mb} + g_{m,QFG}$) helps in increasing the UGBW of the OTA.

To maintain stability of the designed BPA along with aforementioned parameters, R_C has been added in the circuit to obtain the higher values of phase margin as earlier reported in work by Cortes et al. [29].

B. Noise and Offset Voltage Analysis

A combination of transistors (as highlighted for R_{large} in Fig. 2) has been used to obtain such high values of resistance. The input-referred noise voltage of BDQFG transistor is given by Eq. (15).

$$V^2_{Noise,BDQFG} = \left(\left(\frac{g_m}{g_{mb}}\right)^2+\left(\frac{C_{Total}}{C_B}\right)^2\right)V^2_{Noise,GD} \tag{15}$$

$V^2_{Noise,GD}$ is the input-referred noise of Gate driven MOSFET. High values of R_{large} also affect the input-referred noise voltage of BDQFG transistor as per Eq. (15–16). The modified value of input-referred noise voltage is given in the Eq. (16).

$$V^2_{Noise,BDQFG} = \left(\left(\frac{g_m}{g_{mb}}\right)^2+\left(\frac{1}{2\pi R_{large}f_{cut,off}C_B}\right)^2\right)V^2_{Noise,GD} \tag{16}$$

A close look at Eq. (16) leads to the fact that high values of R_{large} help in reducing the input-referred noise voltage. This particular approach has been used here in this work to design an OTA capable of providing lower values of NEF.

Another parameter to design a high quality BPA i.e. lowest possible values of offset voltage have been obtained using Eq. (17). The offset voltage (V_{oS}) of BPA also depends on the value of R_{large} used in the BDQFG transistor as described in (17).

$$V_{oS} = \frac{V_{o,ideal}}{\sqrt{g_{m,BDQFG}(r_{o1}||r_{o3})}} \tag{17}$$

$V_{o,ideal}$ is an ideal offset voltage. Presence of $g_{m,BDQFG}$ in the denominator clearly establishes the fact that higher values of R_{large} would help in obtaining smaller values of V_{oS}.

2.1.2 Two-series connected BDQFG PR

Figure 4a shows two-series connected bulk-driven quasi-floating gate pseudo-resistor which have been connected in the feedback path of the BPA shown in Fig. 1 for implementing f_L. The value of I_{Tune}, V_{Bias} and V_B have been set to 10 nA, 1.6 V and 1.6 V respectively. BDQFG pseudo-resistor has been proposed by Sharma et al. [30] which relies on scaling of gate-source and gate-drain capacitances by ac-coupling factor (α) and (1−α) respectively. BDQFG pseudo-resistor emulates values of 1 TΩ with high degree of linearity. Details of the same have been extensively provided in an earlier reported work [30].

Figure 4b shows the plot of incremental resistance (R_{PR}) and current (I_{PR}) on y_1 and y_2 axes respectively, versus swept voltage (ΔV_{XY}) between two series connected BDQFG PRs (M_a devices) shown in Fig. 4b. As can be clearly observed from the Fig. 4b that there exists a linear relationship between I_{PR} and ΔV_{XY} and value of the I_{PR} is 0.2 pA for swing voltage -1 $\leq \Delta V_{XY} \leq$ 1. The value of the R_{PR} over ΔV_{XY} as calculated from Eq. (18) is found to be 2 TΩ and has been marked as almost a straight line in Fig. 4b.

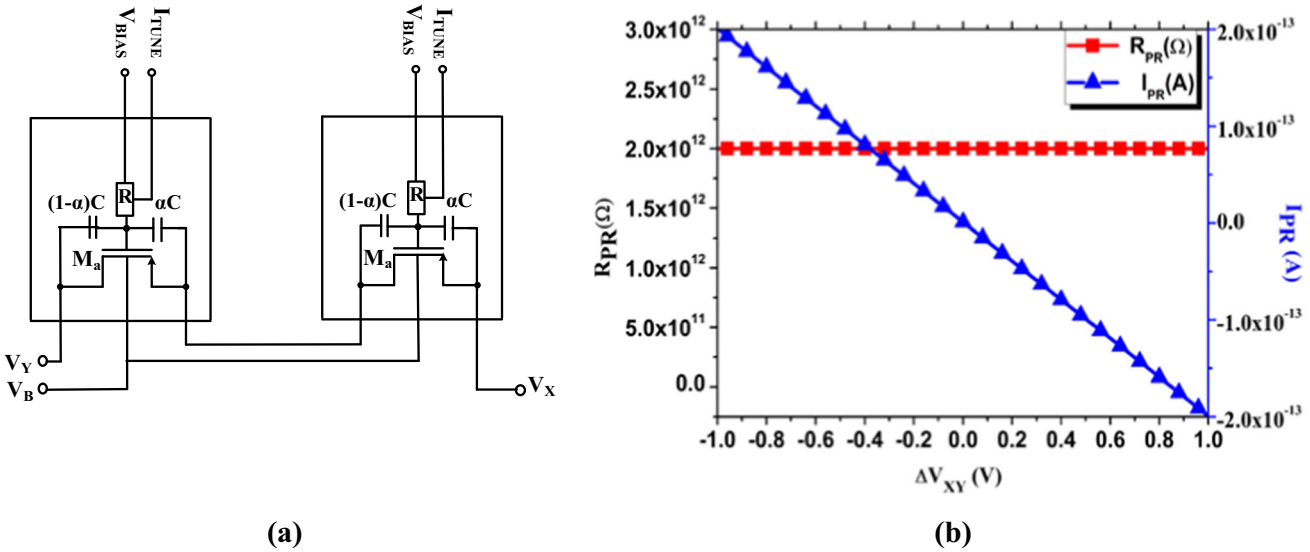
$$R_{PR} = \frac{1}{\delta I_{PR}/\delta V_{XY}} \tag{18}$$

This observed value of R_{PR} is in conformity with theoretical value of 2 TΩ resulting as an outcome of two series connected BDQFG PRs of 1 TΩ each. It is worth mentioning here that while performing simulation of two series connected BDQFG PRs shown in Fig. 4b at 50 Hz, the node V_Y has been kept at 50 μV while the node V_X has been swept from −1 to 1 V. The reason for selection of these values lies in the fact that nodes V_Y and V_X are analogous to input and output terminals of BPA shown in Fig. 1 under Sect. 2. Similar kind of simulation strategy for obtaining V-I and V-R curves for pseudo-resistors has been adopted by Ozalevli et al. [31] and Kassiri et al. [32]. These interpretations convey that such a high magnitude of incremental resistance of the order of 2 TΩ obtained from the combination of two-series connected BDQFG PRs meets the requirement of low f_L of the BPA. This PR topology has been used in feedback path of BPA as indicated by R_F in Fig. 1.

3 Results and discussions

To check the performance of designed BPA circuit shown in Fig. 1, a weak signal of amplitude 100 μV with frequency of 50 Hz has been applied at the input terminal, V_n. As per the mathematical analysis carried out in Sect. 2, a wide range of values of resistance R_{large} have been used in the circuit for purpose of optimization. This has been done so as to get the best possible combination of results in terms of high values of gain and bandwidth along with least possible values of NEF, total-input referred noise voltage and consumed power.

Figure 5a, b show the effect of variations of values of R_{large} on the gain and total input-referred noise. As clear from the Fig. 5a, b, R_{large} have been varied from 10 kΩ to



(a) **(b)**

Fig. 4 **a** Schematic of two-series connected BDQFG PR and **b** simulated I_{PR}, R_{PR} versus V_{XY} for two-series connected BDQFG PR

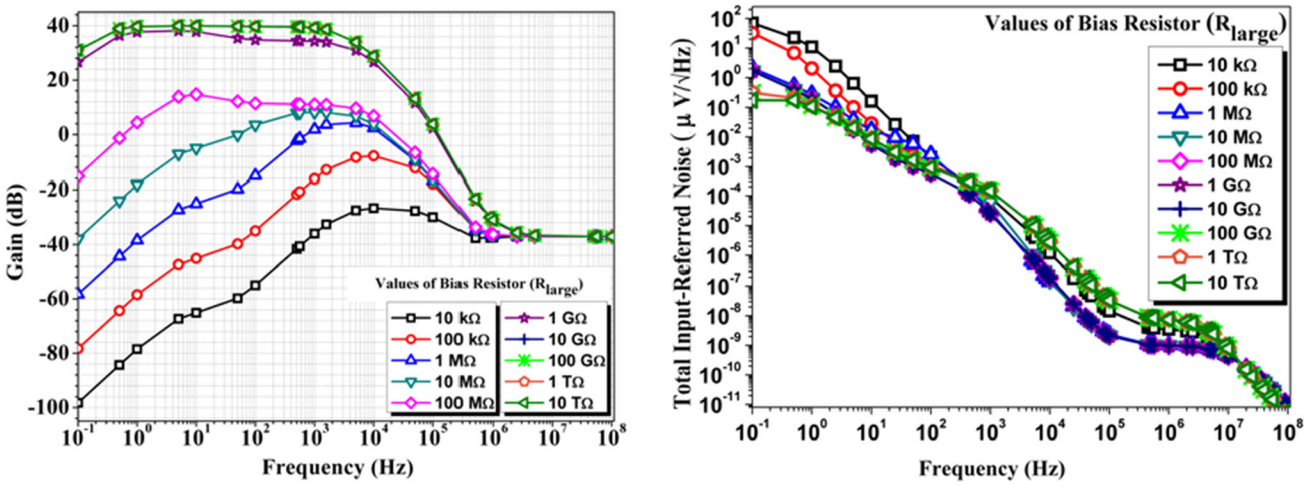


Fig. 5 Simulated plots for biopotential amplifier at different values of bias resistor, R_{large} = 10 kΩ, 100 kΩ, 1 MΩ, 10 MΩ, 100 MΩ, 1 GΩ, 10 GΩ, 100 GΩ, 1 TΩ and 10 TΩ: **a** Transfer function plot. **b** Total input-referred noise (V/$\sqrt{}$Hz) versus frequency

10 TΩ and accordingly gain has been seen to hover between 10 dB to 40 dB over frequency range of f_L of 0.266 Hz to f_{-3dB} of 2.8 kHz for biopotential signal. The values of $V_{in,RMS}$ have been calculated to be 660 mV_{RMS}, 66 mV_{RMS}, 6.860 mV_{RMS}, 0.888 mV_{RMS}, 0.199 mV_{RMS}, 20.2 μV_{RMS}, 59.9 μV_{RMS}, 19.3 μV_{RMS}, 5.015 μV_{RMS} and 3.05 μV_{RMS} at R_{large} of 10 kΩ, 100 kΩ, 1 MΩ, 10 MΩ, 100 MΩ, 1 GΩ, 10 GΩ, 100 GΩ, 1 TΩ and 10 TΩ respectively.

Least values of input-referred noise have been obtained for high values of R_{large} in the range of 1 GΩ to 10 TΩ. Values of gain and noise, thus obtained for high value of R_{large} with the theoretical analysis explained in Sect. 2. Khateb et al. [27] have also projected dependence of gain and noise on the values of R_{large}.

Value of R_{large} = 9.1 TΩ (optimized this way) has been realized using complementary pseudo-resistor structure based on two series connected PMOS and NMOS in the BPA circuit of Fig. 1. Psuedo-Resistor structure has been biased with gate voltages of V_1 = -0.1 V, V_2 = -0.2 V and bulk voltage of transistor M_8 i.e. V_3 = 1.8 V (see Fig. 2). A plot of transfer function and input-referred noise obtained for this value of R_{large} have been shown in Fig. 6a and b respectively. The results show that overall mid-band gain is 39.9 dB from f_L of 0.266 Hz to f_{-3dB} of 2.8 kHz, which is quite close to theoretical values as obtained from Eqs. (1–3). Results have also been obtained for the UGBW values from Fig. 5a. Same were found to be equal to 1.84 kHz, 22.1 kHz, 33.2 kHz, 40.81 kHz, 152.1 kHz, 190.63 kHz, 236.1 kHz, 241.64 kHz, 260.3 kHz and 278.78 kHz at the values of R_{large} of 10 kΩ, 100 kΩ, 1 MΩ, 10 MΩ, 100 MΩ, 1 GΩ, 10 GΩ, 100 GΩ, 1 TΩ and 10 TΩ respectively.

Figure 6b shows the simulated total-input referred noise voltage (V/$\sqrt{}$Hz) versus frequency spectrum. The total input-referred noise voltage ($V_{in,RMS}$) for BPA integrated over bandwidth of 0.266 Hz to 2.8 kHz is evaluated to be 3.15 μV_{RMS}. The reason for such low values of noise is attributed to high value of resistance which can be inferred from Eq. (16).

Also it can be interpreted from Eq. (19) that $V_{in,RMS}$ of the BPA is directly related to the total-input referred noise voltage of OTA ($V_{in,OTA}$) [9].

$$V_{in,RMS} = \left(\frac{2C_{in} + C_B}{C_{in}}\right)^2 V_{in,OTA} \tag{19}$$

The noise efficiency factor (*NEF*) at $R_{large} \approx$ 9.1 TΩ has been obtained to be 4.14 by (20) [9].

$$NEF = V_{in,RMS}\sqrt{\frac{2}{\pi V_T}\left(\frac{I_{Supply}}{4kT}\right)\frac{1}{BW}} \tag{20}$$

I_{Supply} is the supply current; T is the temperature; (300 K); BW is the bandwidth of the biopotential amplifier and k is the Boltzmann constant

Figure 7 shows simulated transient response of the proposed BPA. While Fig. 7a depicts the weak input signal of amplitude 100 μV with the frequency of 50 Hz, Fig. 7b–d represent the amplified output voltage at the different values of R_{large} i.e. 100 MΩ, 100 GΩ and 9.1 TΩ respectively. A high value of DC offset voltage = 494 mV is generated in the amplified output voltage at R_{large} = 100 MΩ. The values of DC offset voltage were found to reduce to 4.54 mV and 455 μV at R_{large} = 100 GΩ and 9.1 TΩ respectively as illustrated in Fig. 7c and Fig. 7d. This is

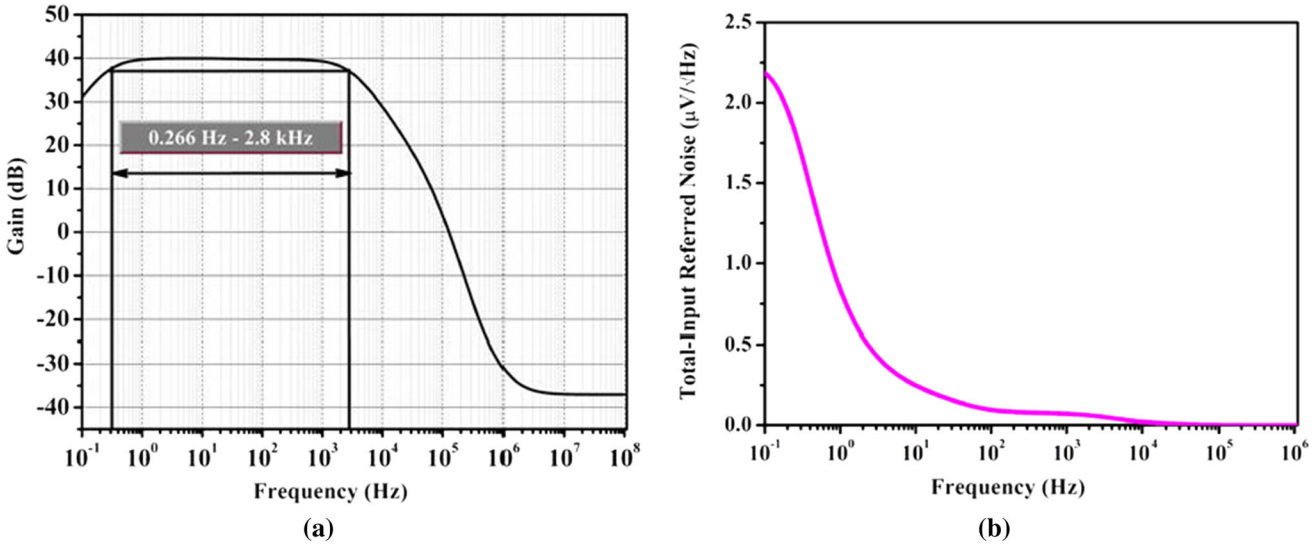


Fig. 6 Simulated plots for biopotential amplifier: at R_{large} = 9.1 TΩ. **a** Transfer function plot and **b** total input-referred noise voltage ($V/\sqrt{Hz}$) versus frequency

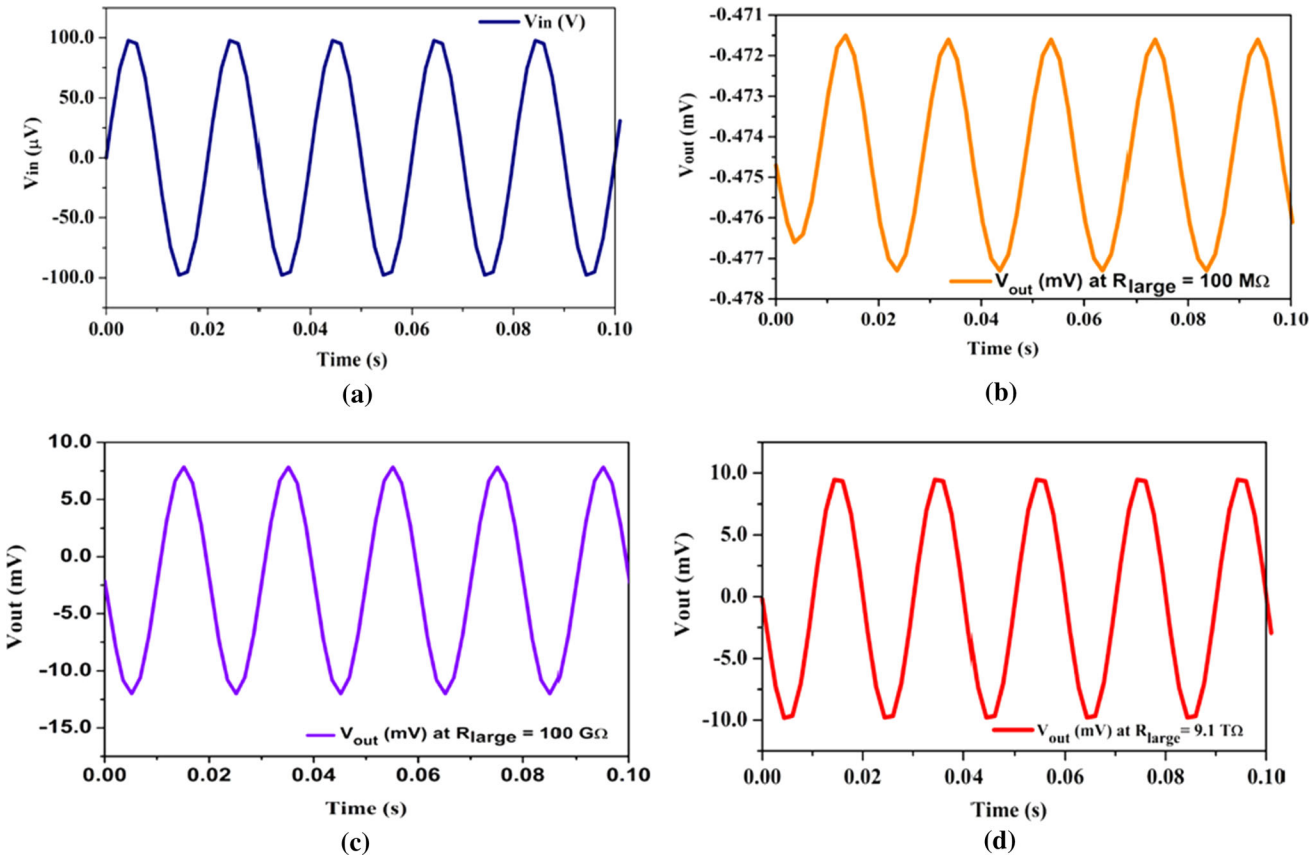


Fig. 7 Simulated transient response for the proposed neural amplifier: **a** input voltage (V_{in}) plot with peak-to-peak voltage ($V_{p\text{-}p}$) = 200 μV at 50 Hz, **b** output voltage plot at R_{large} = 100 MΩ with DC offset voltage of 494 mV, **c** output voltage plot at R_{large} = 100 GΩ with DC offset voltage of 4.54 mV and **d** output voltage plot at R_{large} = 9.1 TΩ with DC offset voltage of 455 μV

as per the mathematical Eq. (17) earlier established in Sect. 2, which clearly revealed the reduction of offset voltage with respective increase in value of R_{large}.

Figure 8a, b show the simulated common-mode gain (A_{CM}) and supply-gain (A_{SG}) versus frequency. From these simulated results, the common-mode gain and supply-gain

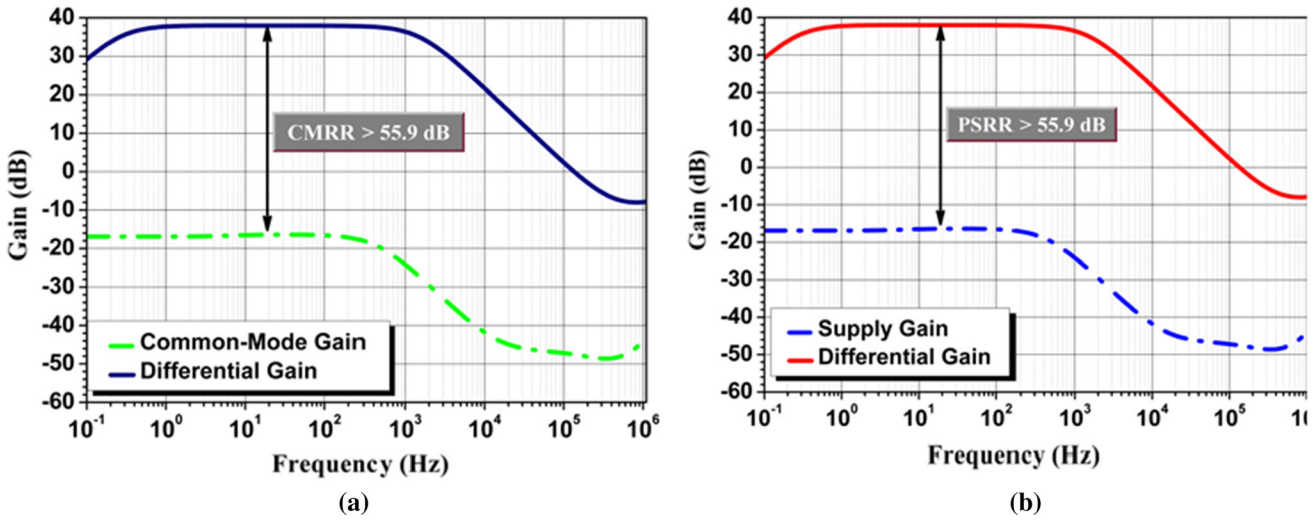


Fig. 8 **a** Common-mode rejection ratio (CMRR) and **b** power-supply rejection ratio (PSRR) for BPA

were found to be -16 dB over bandwidth of 0.266 Hz to 2.8 kHz. With the calculated value of A_{CM} and A_1, CMRR was calculated to be 55.9 dB using Eq. (21).

$$CMRR = 20\log\frac{A_1}{A_{CM}} \tag{21}$$

It can be easily analysed from Eq. 2 that the gain of the BPA depends on the transconductance of OTA i.e. $g_{m,OTA}$. It is also described in Eq. (13), $g_{m,OTA}$ is directly related to the value of R_{large}. Hence, the value of CMRR also depends upon the value of R_{large}. In a similar manner, PSRR also depends on the value of R_{large} as PSRR is also directly related to the gain of BPA described in Eq. (22).

$$PSRR = 20\log\left(\frac{A_1}{A_{SG}}\right) \tag{22}$$

With the calculated value of A_{SG} and A_1, the PSRR was found to be 55 dB using Eq. (22).

For a typical process corner at room temperature (27° C), the phase margin of 65.83 ° is obtained with a gain crossover frequency of 11.851 kHz. Total power consumed by BPA is equal to 1.1 µW for 1.32 µA supply current from a ± 0.5 V voltage supply. Viability of the results obtained this way was also verified over a broader range of process variations.

Table 2 shows the simulation results for different parameters such as power consumption, gain, $V_{in,RMS}$, NEF, bandwidth and offset voltage for typical, fast and slow process variations. As can be clearly seen that the results of most of the parameters remain intact even after subjecting the circuit from slow (Less carrier mobility than conventional carrier mobility) to fast (More carrier mobility than conventional carrier mobility) process variations.

Table 3 shows the comparison of the performance of BPA design presented in this paper along with work of other researchers [9, 11, 33, 34]. It is evident from the Table 3 that the BPA reported by [9] is based on current-mirror topology which achieved gain of 40 dB from 2.2–30 Hz. Usage of power supply of ± 2.5 V along with a supply current of 0.128 µA led to the power consumption of 0.9 µW in this topology. The integrated input-referred noise was reported to be equal to 2.4 μV_{RMS} with the NEF of 6.0. The reported BPA was designed and simulated using 1.5 µm CMOS process. Wattanapanitch et al. [11] demonstrated current-mirror OTA topology based biopotential amplifier which yielded the gain of 40.9 dB over the bandwidth of 0.392–295 Hz. The reported amplifier consumed the power of 2.08 µW from a 2.8 V supply voltage at the supply current of 2.7 µA with the total-input referred noise of 1.66 μV_{RMS} and NEF of 3.21. The proposed amplifier was fabricated in AMI's 0.5 µm CMOS process. Dwivedi et al. [33] demonstrated folded-cascode OTA based neural amplifier which yielded a gain of 31.7 dB over the bandwidth of 3–164 Hz with the very low-power consumption of 0.0684 µW. The topology reported by them operated at the supply voltage of 0.8 V along with the supply current of 0.0854 µA. Although they could obtain

Table 2 Simulation results for process variation

Parameters	Typical	Fast	Slow
Power (µW)	1.1	1.01	1.28
Gain (dB)	39.9	39.8	39.8
$V_{in,RMS}$ (µV)	3.15	3.10	3.25
NEF	4.14	4.12	4.21
Bandwidth (Hz)	0.266–2.8 k	0.257–2.63 k	0.263–2.94 k
Offset Voltage (µV)	455	681	347

Table 3 Comparative analysis of parameters of different neural amplifiers

Year Reference	2003 [9]	2007 [11]	2015 [33]	2018 [34]	This work
Technology (nm)	1500	500	180	180	180
OTA Topology	Current mirror	Current mirror	Folded-cascode	Folded cascode	Miller compensated
V_{dd} (V)	±2.5	2.8	0.8	1.8	±0.5
Supply Current (μA)	0.128	2.7	0.0854	–	1.32
Power (μW)	0.9	2.08	0.0684	3.345	1.1
Gain (dB)	40	40.9	31.7	61.52	39.9
$V_{in,RMS}$ (μV)	2.4	1.66	5.62	3.3	3.15
NEF	6.0	3.21	6.33	2.64	4.14
Bandwidth (Hz)	2.2-30	0.392-295	3-164	13.2 k (-3 dB)	0.266-2.8 k
Application	EEG	LFP	LFP	LFP, Spikes	EEG, LFP, ECoG, ECG

the least value of power consumption, but the reported amplifier provided high values of NEF and total-input referred noise voltage of the order of 6.33 and 5.62 μV_{RMS} respectively. Laskar et al. [34] reported folded-cascode OTA topology based neural amplifier that provided the high gain of 61.52 dB over -3 dB bandwidth of 13.2 kHz, but the reported amplifier consumed the large amount of power of 3.345 μW from a supply voltage of 1.8 V with the total-input referred noise voltage of 3.3 μV_{RMS} and NEF of 2.64.

In comparison to the earlier reported work, the BPA topology presented in this paper achieved gain of 39.9 dB over the bandwidth of 0.266–2.8 kHz. Most remarkably the power consumed is of the order of only 1.1 μW with input-referred noise of 3.15 μV_{RMS} and NEF of 4.14. The reason for low-values of power consumption is based on the fact that BDQFG technique reduces V_T which further helped in cutting down on the requirements of biasing voltage [27]. Though the value of noise is bit higher than the previously reported work by Harrison et al. and Wattanpanitch et al. [9, 11] but is within acceptable range of 1–4 μV_{RMS} as per [26].

4 Conclusion

A biopotential amplifier has been designed and simulated using 0.18 μm technology for an optimum combination of figure of merits like gain (39.9 dB), power consumption (1.1 μW), offset voltage (455 μV), total input-referred noise voltage (3.15 μV_{RMS}) and NEF (4.14). Bulk-Driven Quasi-Floating Gate approach with a special focus on the effect of variation in the values of biasing resistor (R_{large}) has been used in the internal architecture of OTA in the BPA topology worked upon in this paper. Results obtained from detailed mathematical analysis and simulations have been found to be in agreement with each other clearly validating the pre-existing theories. Comparison of the results obtained here with the work reported in the existing literature establishes the potential utility of the BPA circuit (reported in this paper) in the design of more power-efficient biomedical systems. Design of a notch low-pass filter and incorporation of the same in low-frequency based biomedical applications is a further opportunity as an extension of this work.

Acknowledgement Authors would like to thank Vice Chancellor, Chitkara University for all the support to carry out this work and also provide financial support to student authors - Preeti and Kulbhushan. Sharing of Technical knowhow by SCL team members is also deeply acknowledged.

References

1. Chen, W. M., Yang, W. C., Tsai, T. Y., Chiueh, H., & Wu, C. Y. (2011, August). The design of CMOS general-purpose analog front-end circuit with tunable gain and bandwidth for biopotential signal recording systems. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4784–4787). https://doi.org/10.1109/iembs.2011.6091185.
2. Holleman, J. (2016, August). Design considerations for neural amplifiers. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6331–6334). https://doi.org/10.1109/embc.2016.7592176.
3. Mollazadeh, M., Murari, K., Cauwenberghs, G., & Thakor, N. (2008, August). From spikes to EEG: Integrated multichannel and selective acquisition of neuropotentials. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2741–2744). https://doi.org/10.1109/iembs.2008.4649769.
4. Sharma, K., & Sharma, R. (2019). Design considerations for effective neural signal sensing and amplification: a review. *Biomedical Physics & Engineering Express, 5*(4), 1–28. https://doi.org/10.1088/2057-1976/ab1674.

5. Chandrakumar, H., & Marković, D. (2017). A high dynamic-range neural recording chopper amplifier for simultaneous neural recording and stimulation. *IEEE Journal of Solid-State Circuits, 52*(3), 645–656. https://doi.org/10.1109/JSSC.2016.2645611.
6. Wu, C. Y., Cheng, C. H., & Chen, Z. X. (2018). A 16-channel CMOS chopper-stabilized analog front-end ECoG acquisition circuit for a closed-loop epileptic seizure control system. *IEEE Transactions on Biomedical Circuits and Systems, 12*(3), 543–553. https://doi.org/10.1109/TBCAS.2018.2808415.
7. Xu, J., Konijnenburg, M., Ha, H., VanWegberg, R., Song, S., Blanco-Almazan, D., et al. (2018). A 36 µW 11 mm 2 reconfigurable analog front-end for cardiovascular and respiratory signals recording. *IEEE Transactions on Biomedical Circuits and Systems, 12*(4), 774–783. https://doi.org/10.1109/tbcas.2018.2814699.
8. Gosselin, B. (2011). Recent advances in neural recording microsystems. *Sensors, 11*(5), 4572–4597. https://doi.org/10.3390/s110504572.
9. Harrison, R. R., & Charles, C. (2003). A low-power low-noise CMOS amplifier for neural recording applications. *IEEE Journal of Solid-State Circuits, 38*(6), 958–965. https://doi.org/10.1109/JSSC.2003.811979.
10. Mohseni, P., & Najafi, K. (2004). A fully integrated neural recording amplifier with DC input stabilization. *IEEE Transactions on Biomedical Engineering, 51*(5), 832–837. https://doi.org/10.1109/TBME.2004.824126.
11. Wattanapanitch, W., Fee, M., & Sarpeshkar, R. (2007). An energy-efficient micropower neural recording amplifier. *IEEE Transactions on Biomedical Circuits and Systems, 1*(2), 136–147. https://doi.org/10.1109/TBCAS.2007.907868.
12. Gosselin, B., Ayoub, A. E., Roy, J. F., Sawan, M., Lepore, F., Chaudhuri, A., et al. (2009). A mixed-signal multichip neural recording interface with bandwidth reduction. *IEEE Transactions on Biomedical Circuits and Systems, 3*(3), 129–141. https://doi.org/10.1109/TBCAS.2009.2013718.
13. Ng, K. A., & Xu, Y. P. (2013). A compact, low input capacitance neural recording amplifier. *IEEE Transactions on Biomedical Circuits and Systems, 7*(5), 610–620. https://doi.org/10.1109/TBCAS.2013.2280066.
14. Denison, T., Consoer, K., Santa, W., Avestruz, A. T., Cooley, J., & Kelly, A. (2007). A 2 µw, 100 nV/rtHz chopper-stabilized instrumentation amplifier for chronic measurement of neural field potentials. *IEEE Journal of Solid-State Circuits, 42*(12), 2934–2945. https://doi.org/10.1109/JSSC.2007.908664.
15. Ng, K. A., & Chan, P. K. (2005). A CMOS analog front-end IC for portable EEG/ECG monitoring applications. *IEEE Transactions on Circuits and Systems I: Regular Papers, 52*(11), 2335–2347. https://doi.org/10.1109/TCSI.2005.854141.
16. Farshchi, S., Pesterev, A., Nuyujukian, P., Guenterberg, E., Mody, I., & Judy, J. W. (2010). Embedded neural recording with TinyOS-based wireless-enabled processor modules. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 18*(2), 134–141. https://doi.org/10.1109/TNSRE.2009.2039606.
17. Ferreira, L. H., Pimenta, T. C., & Moreno, R. L. (2007). An ultra-low-voltage ultra-low-power CMOS miller OTA with rail-to-rail input/output swing. *IEEE Transactions on Circuits and Systems II: Express Briefs, 54*(10), 843–847. https://doi.org/10.1109/TCSII.2007.902216.
18. Lee, J. Y., & Hwang, S. N. (2008). A high-gain boost converter using voltage-stacking cell. *The Transactions of the Korean Institute of Electrical Engineers, 57*(6), 982–984. https://doi.org/10.1049/el:20080371.
19. Mollazadeh, M., Murari, K., Cauwenberghs, G., & Thakor, N. (2009). Micropower CMOS integrated low-noise amplification, filtering, and digitization of multimodal neuropotentials. *IEEE Transactions on Biomedical Circuits and Systems, 3*(1), 1–10. https://doi.org/10.1109/TBCAS.2008.2005297.
20. Shahrokhi, F., Abdelhalim, K., Serletis, D., Carlen, P. L., & Genov, R. (2010). The 128-channel fully differential digital integrated neural recording and stimulation interface. *IEEE Transactions on Biomedical Circuits and Systems, 4*(3), 149–161. https://doi.org/10.1109/TBCAS.2010.2041350.
21. Wang, T. Y., Liu, L. H., & Peng, S. Y. (2015). A power-efficient highly linear reconfigurable biopotential sensing amplifier using gate-balanced pseudoresistors. *IEEE Transactions on Circuits and Systems II: Express Briefs, 62*(2), 199–203. https://doi.org/10.1109/TCSII.2014.2387685.
22. Ng, K. A., & Xu, Y. P. (2016). A low-power, high CMRR neural amplifier system employing CMOS inverter-based OTAs with CMFB through supply rails. *IEEE Journal of Solid-State Circuits, 51*(3), 724–737. https://doi.org/10.1109/JSSC.2015.2512935.
23. Khateb, F., Dabbous, S. B. A., & Vlassis, S. (2013). A survey of non-conventional techniques for low-voltage low-power analog circuit design. *Radioengineering, 22*(2), 415–427.
24. Khateb, F., Vlassis, S., Kumngern, M., Psychalinos, C., Kulej, T., Vrba, R., et al. (2015). 1 V rectifier based on bulk-driven quasi-floating-gate differential difference amplifiers. *Circuits, Systems, and Signal Processing, 34*(7), 2077–2089. https://doi.org/10.1007/s00034-014-9958-3.
25. Kumngern, M., & Khateb, F. (2016). 0.5 V fully differential current conveyor using bulk-driven quasi-floating-gate technique. *IET Circuits, Devices and Systems, 10*(1), 78–86. https://doi.org/10.1049/iet-cds.2014.0260.
26. Harrison, R. R. (2008). The design of integrated circuits to observe brain activity. *Proceedings of the IEEE, 96*(7), 1203–1216. https://doi.org/10.1109/JPROC.2008.922581.
27. Khateb, F. (2014). Bulk-driven floating-gate and bulk-driven quasi-floating-gate techniques for low-voltage low-power analog circuits design. *AEU-International Journal of Electronics and Communications, 68*(1), 64–72. https://doi.org/10.1016/j.aeue.2013.08.019.
28. Rezaee-Dehsorkh, H., Ravanshad, N., Lotfi, R., Mafinezhad, K., & Sodagar, A. M. (2011). Analysis and design of tunable amplifiers for implantable neural recording applications. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 1*(4), 546–556. https://doi.org/10.1109/JETCAS.2011.2174492.
29. Cortes, F. P., Fabris, E., & Bampi, S. (2004). Analysis and design of amplifiers and comparators in CMOS 0.35 µm technology. *Microelectronics Reliability, 44*(4), 657–664. https://doi.org/10.1016/j.microrel.2003.10.014.
30. Sharma, K., & Sharma, R. (2017). Highly consistent bulk driven quasi floating gate (BDQFG) PMOS pseudo-resistor design and implementation in 0.18 micron meter technology. In *TENCON 2017- IEEE Region 10 Conference* (pp. 488–493). https://doi.org/10.1109/tencon.2017.8227913.
31. Ozalevli, E., & Hasler, P. E. (2008). Tunable highly linear floating-gate CMOS resistor using common-mode linearization technique. *IEEE Transactions on Circuits and Systems I: Regular Papers, 55*(4), 999–1010. https://doi.org/10.1109/TCSI.2008.916459.
32. Kassiri, H., Abdelhalim, K., & Genov, R. (2013, October). Low-distortion super-GOhm subthreshold-MOS resistors for CMOS neural amplifiers. In *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (pp. 270–273). https://doi.org/10.1109/biocas.2013.6679691.
33. Dwivedi, S., & Gogoi, A. K. (2015, February). Local field potential measurement with low-power area-efficient neural recording amplifier. In *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy*

Systems (SPICES) (pp. 1–5). https://doi.org/10.1109/spices.2015.7091507.
34. Laskar, N. M., Guha, K., Nath, S., Chanda, S., Baishnab, K. L., Paul, P. K., et al. (2018). Design of high gain, high bandwidth neural amplifier IC considering noise-power trade-off. *Microsystem Technologies*. https://doi.org/10.1007/s00542-018-4142-5.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Preeti Sharma received her graduation degree with the specialization in Electronics and Communication Engineering from Kurukshetra Institute of Technology and Management, Kurukshetra, India in 2013. She completed her post-graduation from Chitkara University, Punjab, India in July, 2019. Currently, she is working as a doctorate student in Chitkara University, Punjab, India. Her research interests include analog circuit design, semiconductor devices and the implementation of ultra-low power analog front-end for neural recording application.



Kulbhushan Sharma received his B.E. in Electronics and Communication Engineering from Sant Longowal Institute of Engineering and Technology, Punjab, India, in 2011. He is currently working as a doctorate student in Chitkara University, Punjab, India, the same place from where he completed his M.E. degree in 2016. He is working for the implementation of ultra-low-power analog front-end for non-invasive as well implantable neuro-electronics for brain machine interfaces. His current research interests lie in biomedical microsystems, deep brain stimulations and neural prosthetics prototyping.



H.S. Jatana received his engineering education from BITS Pilani. He worked at AMS Austria for porting of SCL CMOS processes at their foundry. He has vast experience on CMOS design, testing, characterization and failure analysis, process development/porting and integration. Presently, He is working as Group-l head at SCL Mohali, India and is responsible for design of VLSI products (ASICS and standard products). His areas of interest are analog design, development of compatible add-on process options compatible to standard CMOS flow and Process Integration issues in DSM era. He is also interested in spreading VLSI education. He has delivered numerous lectures on VLSI at various Institutes.



Jaya Madan received her B.Sc. (H) and M.Sc. degrees in Electronics from the University of Delhi, India, in 2011 and 2013, respectively, and the Ph.D. degree in Applied Physics from Delhi Technological University, New Delhi, India, in 2018. Since 2018, she is working in the VLSI Centre of Excellence, Chitkara University, Punjab, India, where she is currently an Assistant Professor. Her research interests include modeling, designing and simulations of Field effect transistors, and solar cells. She is reviewer of various International journals and conferences. She has authored or coauthored 42 papers published in various international journals and conferences.



Rahul Pandey received his B.Sc. (H) and M.Sc. degrees in Electronics from the University of Delhi, India, in 2011 and 2013, respectively, and the Ph.D. degree in Applied Physics from Delhi Technological University, New Delhi, India, in 2018. Since 2018, he is working in the VLSI Centre of Excellence, Chitkara University, Punjab, India, where he is currently an Assistant Professor. His research interests include modeling, designing and simulations of state of the art solar cells and FETs. He has authored or coauthored 36 papers published in various international journals and conferences. He is reviewer of various International journals and conferences.



Rajnish Sharma is Dean (Research) at CURIN department of Chitkara University. He earned his Ph.D. in 2002 while working in National Physical Laboratory, New Delhi. He has served in Birla Institute of Technology and Science (BITS), Pilani as a Faculty for 6 years. He has more than 40 research papers to his credit in International journals and conferences of repute along with a book titled "Solid State Electronic Devices" from Oxford University Press. He has successfully completed a project sponsored by Ministry of Science and Technology, Government of India on the theme related to the deposition and characterization of Diamond Like Carbon Films for futuristic applications related to microelectronic devices. He is a member of Executive committee of IEEE Delhi section. His current research interests are in the area of modelling of next generation electronic devices, design of neural microsystems and best practices in engineering education.

A comparative analysis of illumination estimation based Image Enhancement techniques

Kavinder Singh
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
Kavinder85@gmail.com

Anil Singh Parihar
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
parihar.anil@gmail.com

2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) 978-1-7281-4142-8/20/$31.00 ©2020 IEEE 10.1109/ic-ETITE47903.2020.195

***Abstract*—Computer vision has gathered the attention of several researchers due to the incremental use of it in almost every field. The performance of computer vision systems is many times limited by input image quality. Thus, image enhancement becomes an essential step in these systems. Retinex based algorithms have proven performance in enhancement of low light images. Many Retinex-based algorithms focus on illumination estimation to perform image enhancement. This paper presents a comparative analysis of state-of-the-art image enhancement algorithms based on illumination estimation. In Retinex based approaches, computation of illumination and reflectance is a challenging task. In early approaches, a smooth image is considered as the illumination. However, in the last decade, various methods for estimation of illumination and reflectance have evolved up to a great extent. In this work, we analyze these image enhancement techniques based on illumination estimation. We perform extensive experimentation on a large set of images with varying illumination. The performance is analyzed both quantitatively and qualitatively. This analysis aims to assist researchers and ignite research to develop new efficient algorithms in this field.**

Keywords—Retinex Theory, Image Enhancement, Illumination, Reflectance, Non-Uniform Illumination

I. INTRODUCTION

Images are the most critical aspect of computer vision. Most of the vision-based techniques (like object detection, object tracking, human activity recognition and many more) require image or video as input to accomplish their task. Vision-based methods require visible images for better performance. Images with low visibility or degraded quality can degrade the performance of vision-based techniques. Approaches are available to enrich the perceivability and nature of an image. Histogram equalization is one of the straightforward method. Histogram equalization in an image attempts to extend the dynamic range to improve the quality. Though, it sometimes prompts halo-effects, over-enhancement, and other limitations. To overcome the limitations of histogram equalization, Dynamic HE [1, 2] based techniques can be used. In DHE there are many sub-histograms to perform HE. DHE based techniques work well for the bright region but unable to perform that well in the darker regions. Other prominent algorithms [3, 4, 5] work well for low contrast images, but again their performance degrades in low light regions. Illumination estimation based image enhancement technique deals with the images that contain regions with non-uniform lighting. Various techniques based on illumination estimation operates at the Retinex image formation model [6]. Retinex model [7, 8] considers the image as the product of reflectance and illumination. Estimation Illumination of an image without prior knowledge of the source of luminance is an arduous task. So many authors presented various approaches depending on their prior knowledge for illumination estimation which is discussed in the next section.

II. RELATED WORK

In our earlier work [9], we discussed various algorithms for image enhancement using the Retinex theory in detail. Recent illumination estimation based enhancement algorithms use refined illumination and reflectance to attain the enhanced result. In this paper we will discuss two types of algorithms, the first type of algorithms computes illumination and to estimate the reflectance component it uses the original image, and the estimated illumination and Another cluster of algorithms compute illumination and reflectance simultaneously from the input image. According to Land *et al.* [10], the Retinex theory can be expressed as:

$$\mathbf{L} = \mathbf{R} \circ \mathbf{I} \tag{1}$$

where $\mathbf{L}$ symbolizes image formed, $\mathbf{R}$ denotes reflectance of the scene captured and $\mathbf{I}$ denotes illumination of the image. Few authors have taken a log of this Retinex model. Illumination estimation from an image without prior knowledge is not a well-posed problem. Jobson *et al.* [11] presented Single Scale Retinex (SSR) which focuses on estimating illumination. In SSR author analyzed the properties of center/surround function for illumination estimation. Then prior is given, Illumination of an image can be considered by the smooth of the image. In another work, Jobson *et al.* [12] presented Multi-Scale Retinex (MSR). In MSR, multiple scales are used to compute the illumination. MSR simultaneously provides Total rendition and color-constancy. Also, Jobson *et al*. [13] presented MSR with color restoration (MSRCR). In MSRCR, color restoration is used to restore the colors in the enhanced image. In MSRCR, color restoration is used to restore the colors in the enhanced image. MSRCR works well, but sometimes result looks unnatural. Wang *et al.* proposed NPEA [14] which computes illumination based on a bright pass filter. NPEA maintains the pairs of pixels in the resultant image like the input image. NPEA fails to provide an appropriate enhancement in low-light images. In Fusion based image enhancement (FBIE) [15], it computes the illumination component by fusing multiple derived components from the input. After computing illumination, it estimates reflectance by the pixel-wise division of input image from estimated illumination. To get the enhanced image, it uses gamma correction on the illumination and then combines the refined illumination with computed reflectance. In LIME [16], the algorithm estimates the illumination map by forming an objective function for optimization. After computing illumination, it uses gamma correction to refine the illumination. With the help of gamma-corrected illumination,

978-1-7281-4142-8/$31.00 ©2020 IEEE
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:12 UTC from IEEE Xplore. Restrictions apply.

it computes the enhanced image. Fu *et al.* [17] presented an approach to estimate reflectance and illumination simultaneously using Probability-based prior (PIE). In another work, Fu *et al.* [18] presented a weighted method over a log-based Retinex model for simultaneous estimation(WIE). Park *et al.* presented a Variational optimization-based algorithm for enhancement using simultaneous estimation (LIVR) [19]. LIVR estimates reflectance and illumination component from the image using adaptive weighted optimization. Li *et al.* proposed SLIE [20] for the enhancement of images with low illumination by estimation of illumination and reflectance component. SLIE used a robust Retinex model to operate.

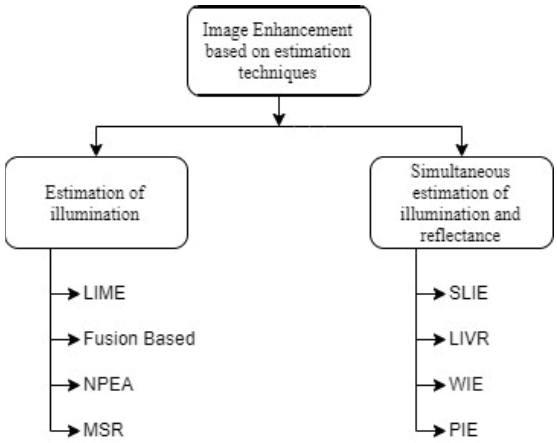


Fig 1. Estimation based Image Enhancement Techniques

Fig 1 shows a few techniques for image enhancement based on estimation methods. Out of these techniques, the paper contains a comparative analysis of a few algorithms out of algorithms shown in Fig 1.

III. METHODOLOGY

MSR performs well for grayscale images but for a color image it sometimes unable to provide proper color in the enhanced image. MSRCR performs well for images, but sometimes results look unnatural. SSR, MSR, MSRCR and few other algorithms based on them consider reflectance as the improved resultant image. However, considering one component (i.e. reflectance) as the enhanced results may be misleading and unnatural. NPEA performs well for images having low variation in lighting. While dealing with images containing high variation in illumination algorithm sometimes fail to provide appropriate enhancement. FBIE performs well for normal lighting images. While dealing with low light images, FBIE produces good results but fails in providing better contrast. LIME performs well while dealing with low light images but it sometimes generates unwanted artifacts in the results. LIME also fails in providing color constancy. PIE fails while dealing with low light images. It computes illumination and reflectance components from the image suitably but fails in providing appropriate enhancement. WIE provides comparably better enhancement than PIE. But while dealing with low light images enhancement is limited. LIVR provides color constancy for normal low light images, but while dealing with dark regions the enhancement is limited. SLIE tries to enhance images by removing noise from the results during enhancement. However, it sometimes considers the small details as noise and removes them from the results which are not suitable in various application of computer vision. After analyzing the algorithms shown in fig 1. FBIE, LIME, WIE, and SLIE are selected for comparison.

A. Fusion-Based Image Enhancement Method With Weak Illumination

Fu. *et al.* presented FBIE [15] for enhancement of image captured under weak lightning. In FBIE authors consider a single image for the multi-fusion process. To perform a multi-fusion algorithm computes various components from a single image based on which algorithm estimates illumination. The algorithm performs morphological closing with a structuring element over the max-channel of the image to compute base input and then normalize it in range 0 to 1. The first component $\mathbf{I}_1$ is the guided smooth [21] of the base input by considering V-channel from HSV model as the guidance image. Second component $\mathbf{I}_2$ computation is performed by applying arc-tan over the base input. The third component $\mathbf{I}_3$ is the "contrast adaptive histogram equalization" [22] of the base input. After the computation of all 3 components, it computes the weights for a particular component. At last, the algorithm performs the multi-fusion operation to compute the illumination over the Gaussian pyramid of the computed weights of the component and Laplacian pyramid of the respective component. Multi-Scale fusion can be expressed as:

$$\mathbf{I}_f = \sum_k G(\mathbf{W}_k) L(\mathbf{I}_k) \tag{2}$$

where $\mathbf{I}_f$ denotes the fusion of various illuminations, $\mathbf{W}_k$ denotes the weights for k[th] illumination scale, $\mathbf{I}_k$ denotes the k[th] illumination scale, $G(\bullet)$ denotes the Gaussian pyramid and $L(\bullet)$ denotes the Laplacian pyramid. The level of the pyramid is taken 6 in the paper [15]. Post computing illumination by multi-fusion, the algorithm computes reflectance by a pixel-wise division of the image with illumination for all color channels. After computing reflectance, the algorithm performs the up-sampling operation over-illumination component to increase the effect of light in weakly lit regions. Finally, it multiplies the up-sampled illumination with computed reflectance for the estimation of the improved resultant image.

B. Lime Via Illumination Map Estimation

Guo *et al.* proposed LIME [16] to improve images with low lighting using the estimation of illumination from the input image. The algorithm considers the max channel as the initial illumination and then uses an objective optimization function to refine the illumination based on a few priors and initial illumination.

$$\min_{\mathbf{I}} \| \hat{\mathbf{I}} - \mathbf{I} \|_F^2 + \alpha \| \mathbf{W} \circ \nabla \mathbf{I} \|_1 \tag{3}$$

where α is a scalar factor and $\mathbf{W}$ is the weight matrix.

The authors provide 2 solutions for the optimization problem. In the first solution named as Exact-Solver, the algorithm uses ADMM [23] with a Lagrangian multiplier to solve the problem. In Exact-Solver, the authors proposed an iterative approach to compute the refined illumination. Exact-Solver stops when it reaches a stopping point which is given in [16]. The algorithm computes the max channel by taking Pixel-wise max from all channels. Then algorithm computes weights using strategy given in [16]. Toeplitz matrix contains the first order derivate of the image in both vertical and horizontal directions. After that, implement the given

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:12 UTC from IEEE Xplore. Restrictions apply.

equations up to the stopping criteria to achieve the results of the Exact-Solver.

In the second approach, termed as Sped-Up-Solver, an alternative approach is proposed to compute refined illumination in lesser time. In Sped-Up-Solver a direct formula is given to compute the refined illumination directly by slight changes in the optimization problem. After the estimation of illumination, it performs gamma correction to improve the visibility of illumination. Also, due to gamma correction final result also contains properties of reflectance as it processes the illumination before the final element-wise division.

C. Simultaneous Estimation Of Reflectance And Illumination Estimation Based On Weighted Variational Model

Fu *et al.* presented WIE [18] for enhancement via estimation of reflectance and illumination simultaneously using an image. WIE operates on a log-based Retinex model. In WIE it considers sum property between log-illumination and log-reflectance to compute both components simultaneously. The major focus of WIE is over finding out the components for image enhancement. The optimization function in WIE consists of the properties additive property between illumination and reflectance of the log-based Retinex model.

$$E(r,i) = \| r + i - l \|_F^2 + \alpha_1 \| e^r \nabla r \|_1 + \alpha_2 \| e^i \nabla i \|_F^2 \quad (4)$$

where r denotes log of reflectance, i denotes log of illumination and l denotes log of image. α_1 and α_2 are positive scalars. WIE computes components based on a given optimization problem.

After the estimation of the components from the image, WIE performs gamma correction over-illumination for improving the visibility of the darker regions.

D. Structure-Revealing Lime Via Robust Retinex Model

Li *et al.* proposed SIE [20] to enhance the image with low lighting. Also, remove noise [24] from the enhanced image. The authors considered a variation of the Retinex model for their approach known as Robust-Retinex. Robust-Retinex considers the image as the sum of noise along with normal Retinex image formation. SLIE can be expressed as:

$$\mathbf{L} = \mathbf{R} \circ \mathbf{I} + \mathbf{N} \quad (5)$$

where **N** denotes the noise in the image. Based on the model authors formulated an objective optimization function to compute the illumination, Reflectance, and noise from the image simultaneously.

$$\min_{R,I,N} \| R \circ I + N - L \|_2^2 + \beta \| \nabla I \|_1 + \gamma \| \nabla R - G \|_2^2 + \omega \| N \|_2^2 \quad (6)$$

where β, γ *and* ω are the positive scalars. G denotes the guidance matrix for reflectance. The algorithm considered prior to reflectance for which authors considered the product of gradient and an exponential gradient-based parameter which also claimed for avoiding the amplification of intensive noise. SLIE is an iterative approach with multiple solutions of sub-problems that stops at the stopping criteria given based on threshold difference between previous and next reflectance or illumination. In the iterative method, the algorithm computes solution for reflectance in which initial reflectance is not required due to which in algorithm reflectance is not having any initial initialization. Then it computes the illumination following which noise computation is done based on computed illumination and reflectance w.r.t original image. After computation of basic parameters, it requires to compute the equality-constraint, Lagrangian-multiplier and the scalars. The algorithm after reaching the stopping point algorithm returns the illumination and reflectance. After which illumination is undergone for gamma-correction to improve the visibility and then it gets multiplied to the reflectance to return the enhanced image.

IV. Results and Performance Analysis

The algorithms used for the comparison are FBIE, LIME, WIE and SIE. The respective authors give the codes for FBIE, WIE, LIME and SLIE on their website. All parameters are selected as per given instructions in respective papers. The analysis of these algorithms is performed over 200 images taken from ExDark Dataset [25], HDR Dataset [26] and Berkeley image dataset [27]. The comparison is performed using quantitative and qualitative analysis.

A. Quantitative Assessment

The Quantitative assessment is required due to a few limitations with the human visual system. The contrast of the image cannot be quantitatively assessed very easily. Sometimes it is possible that human can have on subject-oriented perception. There is various contrast assessment measure available but no measure is present which is universally accepted. We are using 2 measures for the contrast measurement of the images which are Discrete Entropy and LOE.

A statistical measure of randomness is known as the discrete entropy. The greater the value the more information is present in the image and vice versa. Sometimes the value of entropy is affected by the noise i.e., higher noise enhancement leads to the high value of the entropy. DE mathematically can be represented as:

$$H(x) = -\sum_{m=1}^{k} p(m) \log p(m) \quad (7)$$

Where p(m) denotes the probability of mth grayscale. K denotes the number of intensity levels in the image. LOE [14] is an average of the relative order of lightness. The relative order of lightness can be computed by taking an average of exclusive-OR for all pairs in the image where the first term in the pair should be greater than the second term.

TABLE I
Discrete Entropy of sample images

Image	FBIE	LIME	WIE	SLIE
Dog	0.75	0.47	0.67	0.70
Forest	0.58	0.30	0.61	0.56
Building	0.79	0.52	0.67	0.77
Child	0.71	0.41	0.64	0.55
Lake	0.47	0.43	0.43	0.43
Cycle	0.51	0.38	0.58	0.47
River	0.67	0.46	0.63	0.33
Mean	**0.64**	0.42	0.60	0.54

LOE can be represented as:

$$LOE = \frac{1}{m * n} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(U\left(Q(i)Q(j)\right) \oplus U\left(Q_e(i)Q_e(j)\right) \right) \quad (8)$$

Where m and n denote the number of rows and columns respectively, function U(a,b) will be 1 iff a>b otherwise it will be 0. Q(.) denotes pixel-wise max of all channels.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:12 UTC from IEEE Xplore. Restrictions apply.

LOE is a referential measure, but the reference image here is the input image. All methods are improving illumination be gamma correction so larger value here is better. If a proper reference image is taken then the lower value of LOE is better.

TABLE I
LOE of sample images $*10^{-3}$

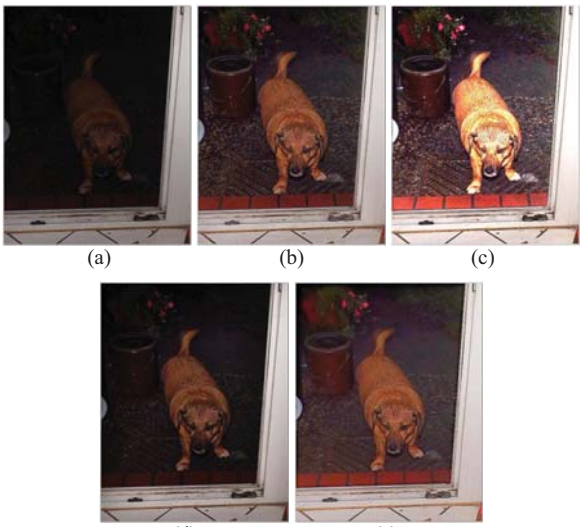
Image	FBIE	LIME	WIE	SLIE
Dog	0.09	0.41	0.21	0.22
Forest	0.53	0.91	0.08	0.51
Building	0.14	0.45	0.15	0.21
Child	0.10	0.27	0.02	0.13
Lake	0.18	0.14	0.06	0.14
Cycle	0.25	0.45	0.13	0.28
River	0.24	0.35	0.14	0.59
Mean	0.22	**0.43**	0.11	0.30

Type your text

B. *Visual Assessment*

The measurement of contrast improvement cannot be determined only by Quantitative assessment. Most of the times, these methods fail in distinguishing between the better contrast & over-enhancement or noise-enhancement. So a visual assessment is required for measuring the quality of output image. Visual assessment is also needed to find out that no artifacts occurred in the result due to the application of the algorithm. The output from the algorithm is given from fig 2-5 for a few sample images. The improved contrast images utilizing distinct methods are compared in this section based on visual representation.

In Fig 2(b)-4(b), it can be noticed that the enhancement achieved by FBIE is looking more natural, but it contains poor contrast in both the images. In Fig 5(b) result of FBIE contains better contrast. In Fig 2(c)-5(c), LIME performing better than FBIE in terms of contrast but the results sometimes look unnatural due to poor color constancy and over-enhancement in a region with normal illumination. In fig 2(d)-5(d), WIE is unable to provide good results for dark regions while in a region with normal lighting the enhancement is limited. In Fig 2(e)-5(e), SLIE provides appropriate enhancement but in some regions, it is a smoothing effect in the resultant image which removes small details of the image and sometimes leads to degraded quality. SLIE provides results with better color constancy but the smoothing effect is undesirable.



(a) (b) (c) (d) (e)

Fig. 2 Dog. (a) Input, (b) FBIE, (c) LIME, (d) WIE, (e) SLIE.



(a) (b) (c) (d) (e)

Fig. 3 Forest. (a) Input, (b) FBIE, (c) LIME, (d) WIE, (e) SLIE.



(a) (b) (c) (d) (e)

Fig. 4 Cycle. (a) Input, (b) FBIE, (c) LIME, (d) WIE, (e) SLIE.



(a) (b) (c) (d) (e)

Fig. 5 River. (a) Input, (b) FBIE, (c) LIME, (d) WIE, (e) SLIE.

C. *Computational Cost*

FBIE and LIME have low time complexity than WIE and SLIE. The method which estimates reflectance and illumination simultaneously using optimization problem takes a higher amount of time than the method which computes illumination only. For images with resolution 424x640, the average time taken by FBIE and LIME is approximately 0.3 seconds. The average time required by WIE and SLIE is more than 2.3 and 4.7 seconds, respectively. SLIE has a large memory requirement as well as it works on images with the help of vectors and sparse matrices. After a limit of image size, SLIE gives out of memory error due to inefficient use of memory.

V. Conclusion

After analyzing FBIE, LIME, WIE, and SLIE, it can be noticed that Each method does have its advantages and limitations. We can conclude that the algorithms available for simultaneous estimation of reflectance and illumination are not suitable for real-time applications. SLIE performs better for noisy images, but it requires better prior for noise estimation as well along with other priors so that it can distinguish between noise and small details. Illumination estimation based algorithms are suitable for real-time applications as they require less than a second for enhancement. However, LIME requires better prior so that it can provide better color constancy and prevent over-enhancement.

After analyzing the estimation based method of image enhancement, it is noticed that these algorithms can be improved by using better prior to the illumination. However, these algorithms work well for images with weak illumination but post-processing over-illumination can be improved. After the improvement in post-processing, these algorithms can work well for multiple tasks simultaneously rather than focusing on a single task.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:12 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan and O. Chae, "A Dynamic Histogram Equalization for Image Contrast Enhancement," *IEEE Trans. Consum. Electron.,* vol. 53, no. 2, pp. 593-600, 2007.

[2] A. S. Parihar and O. P. Verma, "Contrast enhancement using entropy-based dynamic sub-histogram equalisation," *IET Image Processing,* vol. 10, no. 11, pp. 799-808, 2016.

[3] O. P. V. C. K. Anil Singh Parihar, "Fuzzy-contextual contrast enhancement," *IEEE Transactions on Image Processing,* vol. 26, no. 4, pp. 1810-1819, 2017.

[4] A. S. Parihar, "Entropy-Based Adaptive Gamma Correction for Content Preserving Contrast Enhancement," *International Journal of Pure and Applied Mathematics,* vol. 117, no. 20, pp. 887-893, 2017.

[5] A. S. Parihar, "Fuzzy adaptive gamma correction for contrast enhancement," in *International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, India, 2017.

[6] E. H. Land and J. Mccan, "Lightness and retinex theory," *Journal of Optical Society of America,* vol. 61, no. 1, pp. 2032-2040, 1971.

[7] E. H. Land, "Recent advances in retinex theory and some implications for cortical computations," *Proc. Natl. Acad. Sci. U.S.A.,* vol. 80, pp. 5163-5169, 1983.

[8] E. H. Land, "Recent advances in retinex theory," *Vision Research,* vol. 26, no. 1, pp. 7-21, 1986.

[9] A. S. Parihar and K. Singh, "A study on Retinex Based method for Image Enhancement," in *ICISC*, Coimbatore. India, 2018.

[10] L. E. H and M. J, "Lightness and retinex theory," *Journal of Op tical Society of America,* vol. 61, no. 1, pp. 2032-2040, 1971.

[11] D. J. Jobson, Z.-u. Rahman and G. A. Woodell, "Properties and performance of a center/ surround Retinex," *IEEE Transactions on Image Processing,* vol. 6, no. 3, pp. 451 - 462, 1997.

[12] Z.-u. Rahman, D. Jobson and G.A. Woodell, "Multi-scale retinex for color image enhancement," *IEEE International Conference on ImageProcessing,* vol. 3, p. 1003–1006, 1996.

[13] Z.-u. Rahman, D. J. Jobson and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes.," *TIP,* vol. 6, no. 7, pp. 965-976, 1997.

[14] S. Wang, J. Zheng, H.-M. Hu and B. Li, "Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images," *TIP,* vol. 22, no. 9, pp. 3538-3548, 2013.

[15] x. Fu, d. zeng, y. huang, y. liao, x. ding and j. paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Processing,* vol. 129, pp. 82-96, 2016.

[16] X. Guo, Y. Li and H. Ling, "LIME: Low-light Image Enhancement via Illumination Map Estimation," *TIP,* vol. 26, no. 2, pp. 982-993, 2017.

[17] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang and X. Ding, "A Probabilistic Method for Image Enhancement With Simultaneous Illumination and Reflectance Estimation," *IEEE Transactions on Image Processing,* vol. 24, no. 12, pp. 4965-4977, 2015.

[18] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *CVPR*, Las Vegas, Nevada, 2016.

[19] S. Park, S. Yu, B. Moon, S. Ko and J. Paik, "Low-Light Image Enhancement Using Variational Optimization-based Retinex Model," *IEEE Transactions on Consumer Electronics,* vol. 63, no. 2, pp. 178-184, May 2017.

[20] M. Li, J. Liu, W. Yang, X. Sun and Z. Guo, "Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model," *IEEE Transactions on Image Processing,* vol. 27, no. 6, pp. 2828-2841, 2018.

[21] K. He, J. Sun and X. Tang, "Guided Image Filtering," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 35, no. 6, pp. 1397-1409, 2013.

[22] E. Pisano, S. Zong and B. Hemminger, "contrast limited adaptive histogram image processing to improve the detection of simulated spiculations in dense mammograms," *J Digit Imaging,* vol. 11, no. 4, pp. 193-200, 1998.

[23] Y. wang, w. yin and j. zeng, "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization," *Journal of Scientific Computing,* vol. 78, no. 1, pp. 29-63, 2019.

[24] O. P. Verma, A. S. Parihar and M. Hanmandlu, "Edge preserving fuzzy filter for color images," in *International Conference on Computational Intelligence and Communication Networks*, Bhopal, India, 2010.

[25] Y. P. Loh and C. S. Chan, "Getting to Know Low-light Images with The Exclusively Dark Dataset," *Computer Vision and Image Understanding,* vol. 178, pp. 30-42, 2019.

[26] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (Proc. SIGGRAPH Asia),* vol. 35, no. 6, 2016.

[27] D. M. Malik, C. Fowlkes, D. Tal and J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," *Proc. 8th Int'l Conf. Computer Vision,* vol. 2, pp. 416-423, July,2001.

[28] A. S. Parihar, "A study on brain tumor segmentation using convolution neural network," in *International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, India, 2017.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:12 UTC from IEEE Xplore. Restrictions apply.

A Comparative Study of Subject-Dependent and Subject-Independent Strategies for EEG-Based Emotion Recognition using LSTM Network

Debarshi Nath
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
debarshinath94@gmail.com

Anubhav
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
anubhav2901@gmail.com

Mrigank Singh
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
mriganksingh1@gmail.com

Divyashikha Sethia
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
divyashikha@dtu.ac.in

Diksha Kalra
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kalradiksha11@gmail.com

S. Indu
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
s.indu@dce.ac.in

ABSTRACT
This paper addresses the problem of EEG-based emotion recognition and classification and investigates the performance of classifiers for subject-independent and subject-dependent models separately. The results are compared with other classifiers and also with existing work in the concerned domain as well. We perform the experiments on the publicly available DEAP dataset with band power as the feature and classification accuracies are found pertaining to the widely accepted Valence-Arousal Model. The best results were reported by the LSTM model in case of the subject-dependent model with accuracies of 94.69% and 93.13% on valence and arousal scales respectively. SVM performed the best for the subject-independent model with accuracies of 72.19% on valence scale and 71.25% on arousal scale.

CCS Concepts
• Human-centered computing →Human computer interaction (HCI); • Computing methodologies → Machine learning; Artificial intelligence; Feature selection; Classification and regression trees; Machine learning algorithms; Machine learning approaches; Support vector machines; Neural networks.

Keywords
Electroencephalography (EEG); Brain Computer Interface (BCI); Emotion Recognition; Valence-Arousal Model; Long Short-Term Memory network;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ICCDA 2020, March 9–12, 2020, Silicon Valley, CA, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7644-0/20/03…$15.00
https://doi.org/10.1145/3388142.3388167

1. INTRODUCTION
Assessment of emotion is an area that has been dealt with extensively in recent literature in the domain of Brain-Computer Interface. Recent researches have targeted mainly for people suffering from disorders that hinder or alter their ability to express emotions effectively [10]. Different modalities have facilitated the study of emotions like facial expressions, human speech and tone, and physiological signals. One of such modalities is the electroencephalogram (EEG). The human brain comprises of billions of cells, of which a large number are neurons, and others which aid and facilitate the activity of neurons. Whenever any activity occurs, it generates an electrical impulse in the brain due to which thousands of neurons fire in sync. The activations of neurons within our cranium determine emotions and responses. Brain waves are generated by electrical pulses fired in sync from billions of neurons communicating with each other in the nervous system. Brain waves can be detected using electrodes placed on the scalp using invasive, non-invasive or semi-invasive procedures. We study these brain waves in the form of Electroencephalography (EEG). EEG in recent years has become an essential tool with its applications and utility in neuroscience, Brain-Computer Interfaces (BCI's), and commercial applications as well. Analytical tools prevalent in EEG studies frequently use machine learning techniques to uncover relevant information for neural classification and neuroimaging.

Emotion, as easy as it is to feel, is equally challenging to study and measure objectively. Researchers for its objective analysis have pro-posed several models. Most prominent of those is Valence-Arousal Model by Russell [11] that represents emotions on a 2-D circular space where arousal represents the vertical axis, and valence represents the horizontal axis. Various emotions circumscribe the centre of the valence-arousal plane. This centre represents the neutral valence and a medium value for arousal. This model is widely accepted and used extensively in emotion recognition studies. Figure 1 illustrates thevarious ratings of valence and arousal for different emotions.

We also explore other models available for emotion representation. Bradley et al. [2] proposed another model known as Approach and Withdrawal Model or the Vector model. This model is also a 2-D

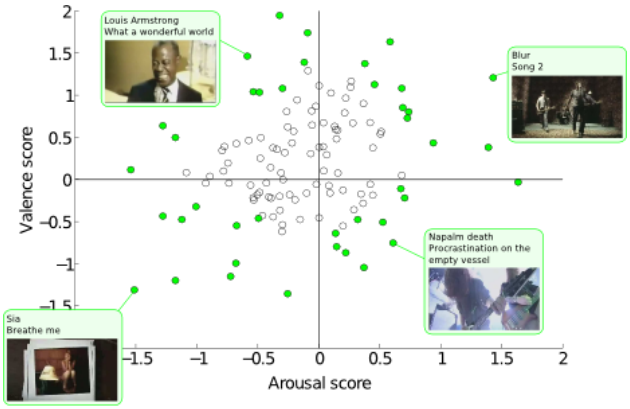


Figure 1. Valence and Arousal ratings corresponding to different emotions as reported in DEAP dataset. [6]

Table 1: Different frequency bands and their significance

Bands	Frequency Range (Hz)	Significance
Delta	0.5 – 4	Deep Sleep, Deepest level of Relaxation
Theta	4 – 8	REM Sleep, Deep and Raw Emotions, Cognitive processing
Alpha	8 – 13	Drowsy state, Relaxation, Calmness
Beta	13 – 30	Conscious state, Thought process
Gamma	> 30	Two different senses at same time

model where the valence determines the direction of emotion. Here, a positive value of valence turns the emotion in the top vector, and on the contrary, the negative value of valence turns the emotion in the down vector. Watson et al. [4] describes a Positive-Negative Model. In this model, the horizontal axis represents low to high negative affect, and the vertical axis represents low to high positive affect.

A typical EEG device measures the electrical activity with the help of electrodes which are in contact with the head of a subject, according to the 10-20 International System. The raw EEG signal is composed of many frequency bands and noise. Hence we need to filter the raw EEG signal then decompose it into constituent frequency bands like Delta (0-4 Hz), Theta (4-8 Hz), Alpha (8-13Hz), Beta (13-30 Hz), and Gamma (>30 Hz) for emotion recognition studies. These bands contain useful information of concern for brain activities. Table 1 describes the significance of these frequency bands. Therefore, EEG signals have broad applications ranging from emotion recognition to diseases and disorders prediction like Sleep Apnea, Epilepsy, and Alzheimer's disease.

2. RELATED WORKS

This section comprises of a detailed review of some prominent researches related to emotion recognition from EEG Signals. For efficient classification of emotions, various factors plays a vital role like pre-processing of brain signals to remove artefacts, feature extraction, feature reduction techniques, and lastly the classifiers used for classification of emotions. Feature extraction plays a vital in emotion recognition as these features represent the information stored in EEG signals. Broadly these features are of three types:

- Time Domain Features
- Frequency Domain Features
- Time-Frequency Domain Features

Jenke et al. [5] have reviewed various feature extraction methods like ReliefF and Min-Redundancy-Max-Relevance (mRMR) for recognition of emotions using EEG signals. Various time-domain features like Event-Related Potentials (ERPs), Statistics of signal such as mean, standard deviation, power, and Hjorth features, Fractal Dimension (FD), Higher-Order Crossings (HOC) are studied. Also, frequency domain features like band power, and Higher-Order Spectra (HOS) as well as time-frequency domain features like Hilbert Huang Spectrum (HHS), Discrete Wavelet Transform (DWT) are described in detail. From our review of the existing literature, we infer that there are two different types of approaches towards emotion recognition algorithm:

- Subject-Dependent
- Subject-Independent

In both the algorithms feature extraction methodology remains the same. However, in the case of subject-dependent on the model algorithm, the classifier is trained for each subject individually, whereas, in the subject-independent model, the classifier is trained for several subjects. We discuss some notable researches related to subject-dependent and subject-independent emotion recognition strategies. Pandey and Seeja [9] used the DEAP Dataset for creating a subject-independent model using wavelet transform as a feature and deep neural network (DNN) as the classifier, achieving a maximum classification accuracy of 62.5% for valance, and 64.25% for arousal. Liu and Sourina [8] proposed a real-time subject dependent emotion recognition technique from EEG Signals using fractal dimension (FD) in combination with statistical and higher Order Crossings (HOC) as feature and SVM as classifier. 14-channel Emotiv device was used for data collection. Sixteen subjects were shown visual stimuli and eight different emotions were recognized. Maximum mean accuracy of 77.81% was achieved for classifying two different emotions. The proposed method is validated on the DEAP Dataset, achieving a mean accuracy of 85.38% for classifying two emotions. Various deep learning applications in Natural Language Processing (NLP) demonstrates the memorizing capability of the LSTM model. Hence, LSTM model can be instrumental in emotion recognition. We also describe other researches employing LSTM model for the task of emotion recognition. Alhagry et al. [1] employed the DEAP dataset for emotion recognition. LSTM model was used for the classification of emotions of 32 subjects individually and achieved 85.65%, 85.45%, and 87.99% accuracy corresponding to arousal, valence, and liking. Li et al. [7] classified human emotions from EEG signals. RASM is used as a feature which is extracted from the DEAP dataset. LSTM network is used as the classifier, achieving a mean accuracy of 76.67% for valence.

Yang et al. [14] employs a parallel combination of Convolutional Neural Network (CNN) and LSTM Network to derive features

from EEG and physiological signals in DEAP dataset. Then, trains the softmax classifier for emotion classification using the subject-dependent strategy, which achieves a mean accuracy of 90.80% and 91.03% for valence and arousal. In this paper, we explore two different training strategies, namely subject-dependent and subject-independent. For both environments, we use band power feature extracted from the raw EEG signals and test different classification techniques like KNN, SVM, Decision Tree, Random Forest, and LSTM model. The next section describes the methodology adopted in this research.

3. METHODOLOGY

3.1 Dataset

The DEAP dataset [6] is a publicly available multimodal dataset designed specifically for emotion analysis. This dataset describes the recordings of EEG and other physiological signals for 32 participants while they watched 40 one-minute-long excerpts of selected videos. A standard EEG headset with 32 channels and sensors like EOG, EMG, and temperature record signals with a sampling frequency of 128 Hz. Then the recorded signals are passed from a bandpass filter to remove noise and artefacts like eye blinks. For assigning an emotion to the signals in terms of valence, arousal, dominance, and liking the participants rated each video with values ranging from 1-9.

3.2 Preprocessing

We observe the raw EEG signals of the DEAP dataset and try to find any similarity between EEG signals of each subject, and also any similarity between different trials for the same subject. We infer that the EEG signal of each individual is unique. There exists no similarity in the EEG signals of different individuals, even if the signals are from the same trial. This is intuitive as each individual possesses different emotional limits and affinities.

However, it is also worth noting that there exist similarities in the EEG signals of the same subject across various trials. From these two observations, we conclude that there is a scope for developing two separate models for studying the emotional states for the subjects, which is also validated by the availability of separate researches into the two models. A subject-dependent model would facilitate the recognition of the emotional state of a particular individual the model is trained on the same individual in prior. A subject-independent model will help determine the emotional state of a person whose previous EEG signals have never been recorded. These are two possible use cases of the two models.

The signals reported in the DEAP data set have a time duration of 63s (3s prior and 60s for the video). Although the 3s prior signal is not affected by the stimulus video, the signals displayed during the stimulus and the final value of emotion may have some correlation with the prior signals. So in this study, we do not remove prior signals before feature extraction. As previous studies suggest, we explore the frequency domain features where we extract the band power of different bands of the EEG signals. We use the Welch method with the Hanning window of 1s to determine Power Spectral Density as displayed in Figure 2. To obtain minute samples of information, we employ a stride of 0.25s for the entire signal. Thus, we obtain 249 band power values for different time instances. We use only the EEG signals for feature extraction and modelling to determine regions of brains responsible for a specific emotion.

Even though the anatomy of brains is similar for every human, but the consciousness of everyone is unique. This fact explains the diversity in the activations of brain, which is validated by the variance in EEG signals. So to train generalized models for the emotion recognition task, we require an extensive database. As the DEAP dataset reports EEG and physiological signals of 32 participants only, it is not currently possible to account for such diversity.

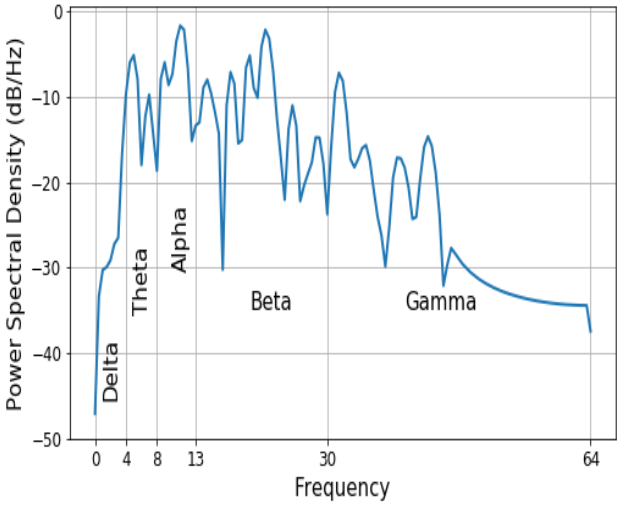


Figure 2. Power Spectral Density of EEG signal

This lack of database motivates to contrast the subject-dependent and subject-independent strategies to train the classifiers for emotion recognition task using EEG signals.

3.3 LSTM Model

In this paper, we test the LSTM network as an efficient tool for the prediction of an individual's emotion. In Natural Language Processing (NLP), the LSTM networks play an influential role in remembering long-term as well as short-term dependencies. The LSTM cell utilizes the following equations for employing the memory retention property.

$$\hat{c}^{<t>} = tanh(\, W_c\, [a^{<t-1>} : x^{<t>}] + b_c\,) \quad (1)$$

$$\Gamma_u = \sigma(W_u\, [a^{<t-1>} : x^{<t>}] + b_u\,) \quad (2)$$

$$\Gamma_f = \sigma(W_f\, [a^{<t-1>} : x^{<t>}] + b_f\,) \quad (3)$$

$$\Gamma_o = \sigma(W_o\, [a^{<t-1>} : x^{<t>}] + b_o\,) \quad (4)$$

$$c^{<t>} = \Gamma_u * \hat{c}^{<t>} + \Gamma_f * c^{<t-1>} \quad (5)$$

$$a^{<t>} = \Gamma_o * tanh(c^{<t>}) \quad (6)$$

Here, the input and output values of the LSTM cell for time t is $x^{<t>}$, and $a^{<t>}$. The symbols W and b represent Weights and Biases of different gates and candidate value for memory. Γ_u,Γ_f, Γ_o represent the Update, Forget, and Output gate for regulating the state of a LSTM cell. $c^{<t>}$ represents the memory of the cell while the $\hat{c}^{<t>}$ represents the candidate value for memory update. The following equations demonstrate the σ(x) and tanh(x) non-linear activation functions:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \quad (7)$$

$$tanh(x) = \frac{e^{x}+e^{-x}}{e^{x}-e^{-x}} \quad (8)$$

Since the previous activations of brain and events critically affect the subsequent activations of the brain, thus remembering past events may enable the LSTM model to gain insights from the patterns of EEG signals. Hence, we test the LSTM model for the emotion recognition task using the EEG signals. Figure 3 displays

the best LSTM model configuration for subject-dependent training strategy. To ensure GPU support, we implement the LSTM

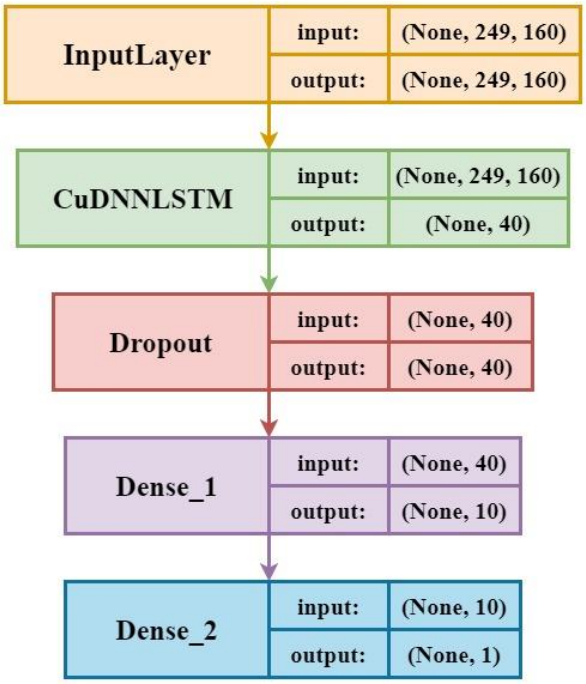


Figure 3. Proposed LSTM Model

model in Python 3 using Keras library with Tensorflow backend on the Google Colab platform. We experiment with several configurations of following LSTM model parameters:

- Number of LSTM Nodes: 32, 40
- Number of Dense 1 Nodes: 10, 20, 24, 32, 40
- Activation for Dense 1: 'relu', 'sigmoid', 'tanh'
- Learning Rate: 1, 1e-1, 1e-2, 1e-3, 1e-4
- Learning Rate Decay: 1e-5, 1e-6

We get the best testing accuracies for the model having 40 nodes in LSTM layer, 10 nodes in Dense 1 layer with 'tanh' activation function, and a single node with 'sigmoid' activation function in Dense 2 layer. To avoid over-fitting, we use 25% dropout between the LSTM Layer and Dense 1 Layer. This model minimizes the log-loss function using the Stochastic Gradient Descent (SGD) optimization algorithm. Following equation displays log-loss function:

$$L(y,\ \hat{y}) = -y\ log(\hat{y}) - (1-y)\ log(1-\hat{y}) \qquad (9)$$

Best results are obtained with the configuration of SGD with learning rate=$1 * e^{-2}$, learning rate decay constant=$1 * e^{-5}$, and momentum constant= 0.9.

4. RESULTS

To contrast the subject-dependent and subject-independent strategies, we examine the performances of different classifiers such as KNN, SVM, Decision Tree, Random Forest, and LSTM model. We train the LSTM model, and other classifiers on the similarly pre-processed dataset having band power features extracted from the EEG signals. The DEAP dataset reports the values of Valence and Arousal on a scale of 1-9; we use 5 as the threshold value for categorizing the low and high values. For determining the accuracy in prediction, we convert the predicted values of different classifiers into labels 0 (low) and 1 (high). Thus,

Table 2. Testing accuracies for Subject-Dependent and Subject-Independent models

Models	Subject-Dependent		Subject-Independent	
	Valence	Arousal	Valence	Arousal
KNN	86.03	79.64	70.86	68.36
SVM	76.56	72.66	**72.19**	**71.25**
Decision Tree	71.10	67.97	58.13	55.63
Random Forest	81.25	81.19	61.95	61.25
LSTM	**94.69**	**93.13**	70.31	69.53

we calculate the prediction accuracy using the equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (10)$$

Here, we determine *True Positive (TP), True Negative (TN), False Negative (FN), and False Positive (FP)* values from the confusion matrix. We follow both the subject-dependent and subject-independent strategies by separately training and testing the classifiers. We test these classifiers several times to ensure that the results observed are significant. Table 2 highlights the average testing accuracies for both the subject-dependent and subject-independent training strategies.

On analyzing the results, we observe that for the subject-dependent strategy, the LSTM model outperforms the other classifiers by a large margin. Whereas in the subject-independent strategy, the SVM model performs better. In subject-dependent conditions, we observe a significant increment of about 16% and 18% for valence and arousal scales, respectively when comparing the performance of LSTM model with other classifiers. We obtain a maximum improvement in average testing accuracy of 18% for valence scale and 20% for arousal scale when comparing the SVM classifier with the LSTM model.

We also contrast our results with the results obtained by Yang et al. [14] following a similar experimental procedure with a parallel combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) model. Here, we observe a gain of 4% for valence and 2% for arousal in average testing accuracies for all the subjects. Proposed LSTM model for subject-dependent approach observes a significant increment of 9% and 7.5% in valence and arousal when compared with Alhagry et al. [1] and 14% and 19% in valence and arousal when compared with Xing et al. [13]. Figure 4 describes the average testing accuracies with positive and negative deviations for valence and arousal.

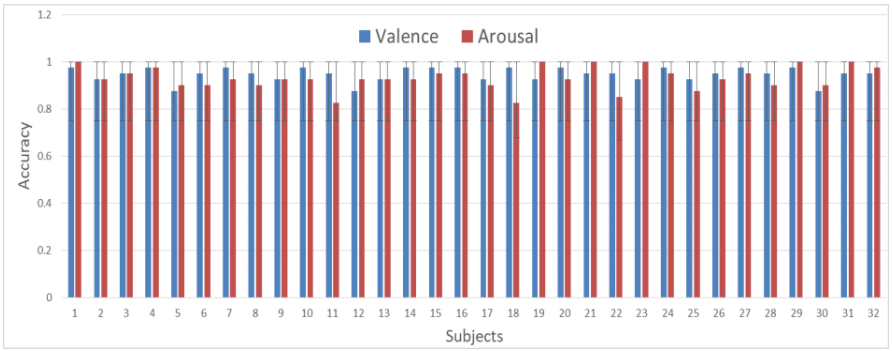


Figure 4. Testing accuracies for 32 subjects using LSTM model for subject-dependent model

For subject-independent conditions, we observe a drop in prediction accuracy of all classifiers. Nevertheless, the performance of the LSTM model drops most significantly, i.e. almost 24% for valence and arousal. This drop in performance indicates that the LSTM model requires an extensive data set for achieving generalization. For the subject-independent model, the best results were achieved by SVM, showing better performance by approximately 6% on the valence scale and 7% on the arousal scale over other classifiers.

5. CONCLUSION

In this work, we evaluate power spectral density over the 32 channels of EEG from the DEAP dataset. We decompose them into five bands of frequencies, Delta, Theta, Alpha, Beta, and Gamma to derive the band power of each band of frequency. We use this band power over each trial as a feature for classifying valence and arousal rating of the subject. We evaluate the performance of classification using KNN, SVM, Decision Tree, Random Forest, and LSTM as our classifiers and compare our results. We find an average increment of 16% for valence and 17% for arousal in the average testing accuracies for the subject dependent model by LSTM model over other classifiers. The maximum classification accuracy of 94.69% for valence and 93.13% for arousal was achieved using the LSTM classifier, which outperforms the state-of-the-art classifiers. For the subject-independent model, the best results were achieved by SVM, showing an average increment of approximately 6% on the valence scale and 7% on the arousal scale over other classifiers. Thus, to create a generalized Deep Learning models for excellent performance on emotion recognition, we require an extensive database.

For further work, we can extend the proposed experimental setup for real-time applications. For improvements further, we can add more features from EEG channels along with physiological features and test for their performance. A subset of the channels for feature generation, rather than using all the channels of EEG, may perform better in terms of accuracy as demonstrated by the works of Wichakam and Vateekul [12]. Other emotion models beyond the Valence-Arousal Model can also be addressed using the strategy used in this paper, like the 3-D emotion model used in the work of Dabas et al. [3].

6. REFERENCES

[1] Salma Alhagry, Aly Aly Fahmy, and Reda A El-Khoribi. 2017. Emotion Recognition based on EEG using LSTM Recurrent Neural Network.Emotion8, 10 (2017), 355–358.

[2] Margaret M Bradley, Mark K Greenwald, Margaret C Petry, and Peter J Lang. 1992.Remembering pictures: Pleasure and arousal in memory. Journal of experimental psychology: Learning, Memory, and Cognition18, 2 (1992), 379–390.

[3] Harsh Dabas, Chaitanya Sethi, Chirag Dua, Mohit Dalawat, and Divyashikha Sethia. 2018. Emotion Classification Using EEG Signals. In Proc. ACM Int. Conf. Computer Science and Artificial Intelligence. ACM, 380–384.

[4] Watson David and Tellegen Auke. 1985. Toward a consensual structure of mood. Psychological bulletin 98, 2 (1985), 219–235.

[5] Robert Jenke, Angelika Peer, and Martin Buss. 2014. Feature Extraction and Selection for Emotion Recognition from EEG.IEEE Trans. on Affective Computing 5, 3 (2014), 327–339.

[6] Sander Koelstra et al.2011. Deap: A Database for Emotion Analysis; Using Physiological Signals. IEEE Trans. Affective Computing 3, 1 (2011), 18–31.

[7] Zhenqi Li, Xiang Tian, Lin Shu, Xiangmin Xu, and Bin Hu. 2017. Emotion Recognition from EEG using RASM and LSTM. In International Conference on Internet Multimedia Computing and Service. Springer, 310–318.

[8] Yisi Liu and Olga Sourina. 2014. Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm. Springer Berlin Heidelberg, Berlin, Heidelberg, 199–223.

[9] Pallavi Pandey and K.R. Seeja. 2019. Subject independent emotion recognition from EEG using VMD and deep learning. Journal of King Saud University -Computer and Information Sciences (2019).

[10] Robert Plutchik. 1991. The Emotions. University Press of America.

[11] James A Russell. 1980. A circumplex model of affect. Journal of personality and social psychology39, 6 (1980), 1161–1178.

[12] Itsara Wichakam and Peerapon Vateekul. 2014. An evaluation of feature extraction in EEG-based emotion prediction with support vector machines. In Proc. IEEE Int. Conf. Joint conference on computer science and software engineering. IEEE, 106–110.

[13] Xiaofen Xing, Zhenqi Li, Tianyuan Xu, Lin Shu, Bin Hu, and Xiangmin Xu. 2019. SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG. Frontiers in Neurorobotics13 (2019), 37.

[14] Yilong Yang, Qingfeng Wu, Ming Qiu, Yingdong Wang, and Xiaowei Chen. 2018. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent

Neural Network. Proc. IEEE Int. Conf. Joint Conference on Neural Networks(2018)



A Multi-task Approach to Open Domain Suggestion Mining Using Language Model for Text Over-Sampling

Maitree Leekha(✉), Mononito Goswami, and Minni Jain

Delhi Technological University, New Delhi, India
{maitreeleekha_bt2k16,mononito_bt2k16,minnijain}@dtu.ac.in

Abstract. Consumer reviews online may contain suggestions useful for improving commercial products and services. Mining suggestions is challenging due to the absence of large labeled and balanced datasets. Furthermore, most prior studies attempting to mine suggestions, have focused on a single domain such as `Hotel` or `Travel` only. In this work, we introduce a novel over-sampling technique to address the problem of class imbalance, and propose a multi-task deep learning approach for mining suggestions from multiple domains. Experimental results on a publicly available dataset show that our over-sampling technique, coupled with the multi-task framework outperforms state-of-the-art open domain suggestion mining models in terms of the F-1 measure and AUC.

Keywords: Open domain suggestion mining · Multi-task learning · Over-sampling techniques · Deep learning

1 Introduction

Consumers often express their opinions towards products and services through online reviews and discussion forums. These reviews may include useful suggestions that can help companies better understand consumer needs and improve their products and services. However, manually mining *suggestions* amid vast numbers of *non-suggestions* can be cumbersome, and equated to finding needles in a haystack. Therefore, designing systems that can automatically mine suggestions is essential. The recent *SemEval* [6] challenge on Suggestion Mining saw many researchers using different techniques to tackle the domain-specific task (*in-domain Suggestion Mining*). However, *open-domain suggestion mining*, which obviates the need for developing separate suggestion mining systems for different domains, is still an emerging research problem. We formally define the problem of open-domain suggestion mining as follows:

Definition 1 (Open-domain Suggestion Mining). *Given a set of reviews* $\mathcal{R} = \{r_1, r_2 \ldots r_n\}$ *from multiple domains in* $\mathcal{D} = d_1 \cup d_2 \cup \ldots d_m$*, train a*

M. Leekha, M. Goswami and M. Jain—Contributed equally and would like to be consider as joint first authors.

© Springer Nature Switzerland AG 2020
J. M. Jose et al. (Eds.): ECIR 2020, LNCS 12036, pp. 223–229, 2020.
https://doi.org/10.1007/978-3-030-45442-5_28

classifier C using $\mathcal{D}$ to predict the nature $n_i \in$ {'suggestion', 'non-suggestion'} of each review r_i.

Building on the work of [5], we design a framework to detect suggestions from multiple domains. We formulate a multitask classification problem to identify both the domain and nature (*suggestion* or *non-suggestion*) of reviews. Furthermore, we also propose a novel language model-based text over-sampling approach to address the class imbalance problem.

2 Methodology

2.1 Dataset and Pre-processing

We use the first publicly available and annotated dataset for suggestion mining from multiple domains created by [5]. It comprises of reviews from four domains namely, `hotel`, `electronics`, `travel` and `software`. During pre-processing, we remove all URLs (eg. *https:// ...*) and punctuation marks, convert the reviews to lower case and lemmatize them. We also pad the text with start `S` and end `E` symbols for over-sampling.

2.2 Over-Sampling Using Language Model: LMOTE

One of the major challenges in mining suggestions is the imbalanced distribution of classes, *i.e.* the number of non-suggestions greatly outweigh the number of suggestions (refer Table 1). To this end, studies frequently utilize *Synthetic Minority Over-sampling Technique* (SMOTE) [1] to over-sample the minority class samples using the text embeddings as features. However, SMOTE works in

Table 1. Datasets and their sources used in our study [5]. The class ratio column highlights the extent of class imbalance in the datasets. The `travel` datasets have lower inter-annotator agreement than the rest, indicating that they may contain confusing reviews which are hard to confidently classify as suggestions or non-suggestions. This also reflects in our classification results.

Dataset	Source	Class ratio (suggestion: non-suggestion)	Inter-annotator agreement
`Hotel train`	Tripadvisor	$^{448}/_{7086} \approx$ 6:100	0.86
`Hotel test`	Tripadvisor	$^{404}/_{3000} \approx$ 13:100	0.86
`Electronics train`	Amazon	$^{324}/_{3458} \approx$ 9:100	0.83
`Electronics test`	Amazon	$^{101}/_{1070} \approx$ 9:100	0.83
`Travel train`	Insight vacations, Fodors	$^{1314}/_{3869} \approx$ 34:100	0.72
`Travel test`	Fodors	$^{229}/_{871} \approx$ 26:100	0.72
`Software train`	Uservoice suggestion forum	$^{1428}/_{4296} \approx$ 33:100	0.81
`Software test`	Uservoice suggestion forum	$^{296}/_{742} \approx$ 39:100	0.81

Table 2. Most frequent 5-grams and their corresponding suggestions sampled using LMOTE. While the suggestions as a whole may not be grammatically correct, their constituent phrases are nevertheless semantically sensible.

Domain	Most frequent 5-gram	A suggestion sampled using LMOTE
`Hotel`	I would definitely recommend hotel	I would definitely recommend hotel good value full ocean view great food worth
`Electronics`	Suggestion get lens protector help	Suggestion get lens protector help protect long lens coating uv 52 lens last long must try
`Travel`	Tipping remember shape luggage concerned	Tipping remember shape luggage concerned heavy luggage rough adviced wheeled duffle wont heavy
`Software`	It would be good if oversight	It would be good if oversight bixby developed bug feels wide back content zoom should be an option

the euclidean space and therefore does not allow an intuitive understanding and representation of the over-sampled data, which is essential for qualitative and error analysis of the classification models. We introduce a novel over-sampling technique, **L**anguage **M**odel-based **O**ver-sampling **Te**chnique (LMOTE), exclusively for text data and note comparable (and even slightly better sometimes) performance to SMOTE. We use LMOTE to over-sample the number of suggestions before training our classification model. For each domain, LMOTE uses the following procedure to over-sample suggestions:

Find Top η `n`-Grams: From all reviews labelled as suggestions (positive samples), sample the top $\eta = 100$ most frequently occurring `n`-grams ($\mathtt{n} = 5$). For example, the phrase *"nice to be able to"* occurred frequently in many domains.

Train Language Model on Positive Samples: Train a BiLSTM language model on the positive samples (suggestions). The BiLSTM model predicts the probability distribution of the next word (w_t) over the whole vocabulary ($V \cup \mathtt{E}$) based on the last $\mathtt{n} = 5$ words ($w_{t-5}, \ldots, w_{t-1}$), *i.e.*, the model learns to predict the probability distribution $P(w_i \mid w_{t-5}\ w_{t-4}\ w_{t-3}\ w_{t-2}\ w_{t-1})\ \forall i \in (V \cup \mathtt{E})$, such that $w_t = \arg\max_{w_i} P(w_i \mid w_{t-5}\ w_{t-4}\ w_{t-3}\ w_{t-2}\ w_{t-1})$.

Generate Synthetic Text Using Language Model and Frequent `n`-Grams: Using the language model and a randomly chosen frequent 5-gram as the seed, we generate text by repeatedly predicting the most probable next word (w_t), until the end symbol `E` is predicted.

Table 2 comprises of the most frequent 5-grams and their corresponding suggestions 'sampled' using LMOTE. In our study, we generate synthetic positive reviews till the number of suggestion and non-suggestion class samples becomes equal in the training set.

Algorithm 1. Language Model-based Over-sampling Technique (LMOTE)

Input:
$\mathcal{D}_{sugg} = \{r_i \in \mathcal{D} \mid n_i = \{\text{`}suggestion\text{'}\}\}$ Suggestions from a particular domain $\mathcal{D}$;
η- Number n-grams to use in LMOTE;
n- type of n-grams *e.g.* 2 for bi-grams, etc.;
$\mathcal{N}$- Number of suggestion samples required.
Output: $\mathcal{S} : \mathcal{N}$ over-sampled suggestions

$n_grams \leftarrow$ **NGrams**$(\mathcal{D}_{sugg}, \eta, n)$
$language_model \leftarrow$ **TrainLanguageModel**$(\mathcal{D}_{sugg}, n)$
Initialize $\mathcal{S} \leftarrow \mathcal{D}_{sugg}$
while $|\mathcal{S}| < \mathcal{N}$ **do**
    $seed \leftarrow$ **random**(n_grams)
    $sample \leftarrow$ **LMOTEGenerate**$(language_model, seed)$
    $\mathcal{S} \leftarrow \mathcal{S} \cup sample$
end while
return $\mathcal{S}$

Algorithm 1 summarizes the LMOTE over-sampling methodology. Following is a brief description of the sub-procedures used in the algorithm:

- **NGrams**$(\mathcal{D}_{sugg}, \eta, n)$**:** It returns the top η *n-grams* from the set of suggestions, D_{sugg}.
- **TrainLanguageModel**$(\mathcal{D}_{sugg}, n)$**:** This procedure trains an *n-gram* BiLSTM Language Model on D_{sugg}.
- **random**(n_grams)- Randomly selects an *n-gram* from the input set.
- **LMOTEGenerate**$(language_model, seed)$**:** The procedure takes as input the trained language model and a randomly chosen *n-gram* from the set of top η *n-grams* as *seed*, and starts generating a review till the end tag, `E` is produced. The procedure is repeated until we have a total of $\mathcal{N}$ suggestion reviews.

2.3 Mining Suggestion Using Multi-task Learning

Multi-task learning (MTL) has been successful in many applications of machine learning since sharing representations between auxiliary tasks allows models to generalize better on the primary task. Figure 1B illustrates 3-dimensional UMAP [4] visualization of *text embeddings* of suggestions, coloured by their domain. These embeddings are outputs of the penultimate layer (dense layer before the final softmax layer) of the *Single task* (STL) ensemble baseline. It can be clearly seen that suggestions from different domains may have varying feature representations. Therefore, we hypothesize that we can identify suggestions better by leveraging domain-specific information using MTL. Therefore, in the MTL setting, given a review r_i in the dataset, D, we aim to identify both the domain of the review, as well as its nature.

2.4 Classification Model

We use an ensemble of three architectures namely, CNN [2] to mirror the spatial perspective and preserve the n-gram representations; Attention Network to learn

the most important features automatically; and a BiLSTM-based text RCNN [3] model to capture the context of a text sequence (Fig. 2). In the MTL setting, the ensemble has two output softmax layers, to predict the domain and nature of a review. The STL baselines on the contrary, only have a singe softmax layer to predict the nature of the review. We use ELMo [7] word embeddings trained on the dataset, as input to the models.

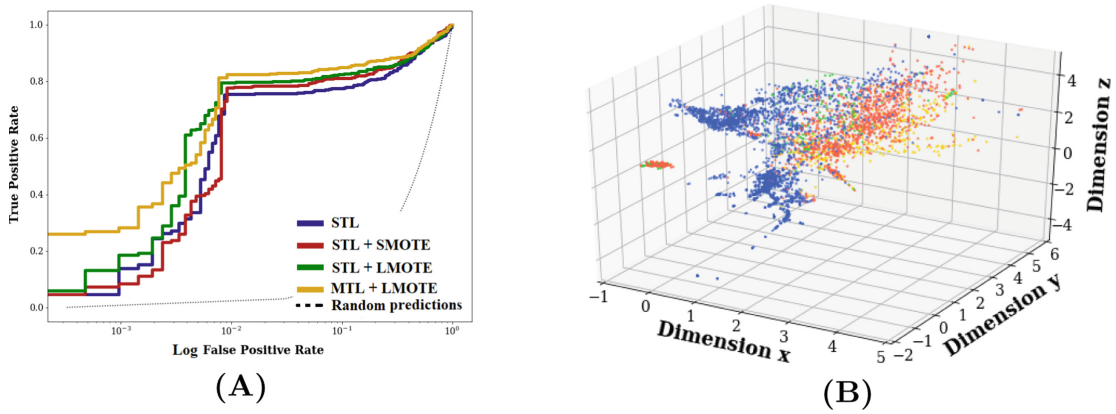


Fig. 1. (A) Receiver operating characteristics (TPR vs. Log FPR) curve pooled across all domains for all models used in this work demonstrates that LMOTE coupled with our multi-task model outperforms other considered alternatives across domains (B) 3-dimensional UMAP visualization of text embeddings of *suggestions* coloured by domain. Suggestions from different domains have distinct feature representations.

3 Results and Discussion

We conducted experiments to assess the impact of over-sampling, the performance of LMOTE and the multi-task model. We used the same train-test split as provided in the dataset for our experiments. All comparisons have been made in terms of the F-1 score of the suggestion class for a fair comparison with prior work on representational learning for open domain suggestion mining [5] (refer *Baseline* in Table 3). For a more insightful evaluation, we also compute the Area under Receiver Operating Characteristic (ROC) curves for all models used in this work. Tables 3, 4 and Figs. 3 and 1A summarize the results of our experiments, and there are several interesting findings:

Over-Sampling Improves Performance. To examine the impact of over-sampling, we compared the performance of our ensemble classifier with and without over-sampling *i.e.* we compared results under the *STL*, *STL + SMOTE* and *STL + LMOTE* columns. Our results confirm that in general, over-sampling suggestions to obtain a balanced dataset improves the performance (F-1 score & AUC) of our classifiers.

LMOTE Performs Comparably to SMOTE. We compared the performance of SMOTE and LMOTE in the single task settings (*STL + SMOTE*

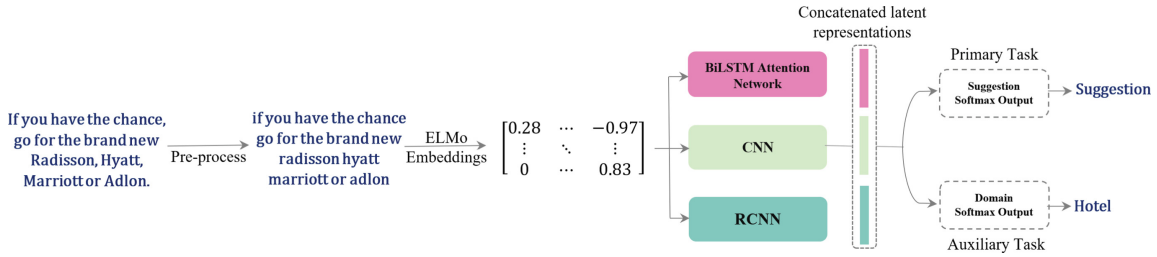


Fig. 2. Our multi-task classification model which consists of an ensemble of RCNN, CNN and BiLSTM attention network. The primary task is predicting the nature of a review (suggestion), while the auxiliary task involves predicting its domain (`hotel`).

and *STL + LMOTE*) and found that LMOTE performs comparably to SMOTE (and even outperforms it in the `electronics` and `software` domains). LMOTE also has the added advantage of resulting in intelligible samples which can be used to qualitatively analyze and troubleshoot deep learning based systems. For instance, consider suggestions *created* by LMOTE in Table 2. While the suggestions may not be grammatically correct, their constituent phrases are nevertheless semantically sensible.

Table 3. Performance evaluation using F-1 score. Multi-task learning with LMOTE outperforms other alternatives in open-domain suggestion mining. Furthermore, owing to potentially confusing reviews in the travel domain (Table 1), its F-1 scores are significantly lower than the other domains.

Domain	Baseline	STL	STL+SMOTE	STL+LMOTE	MTL+LMOTE
`Hotel`	0.77 (LSTM)	0.79	0.83	0.83	**0.86**
`Electronics`	0.78 (SVM)	0.80	0.80	0.83	**0.83**
`Travel`	0.66 (SVM)	0.65	0.68	0.69	**0.71**
`Software`	0.80 (LSTM)	0.79	0.81	0.84	**0.88**

Table 4. Performance evaluation using area under ROC with 95% confidence intervals. Multi-task learning with LMOTE outperforms other alternatives in open-domain suggestion mining. Multi-task learning leads to a significant improvement in AUC over its single task counterpart. (AUCs for baseline models proposed by [5] were unavailable.)

Domain	STL	STL+SMOTE	STL+LMOTE	MTL+LMOTE
`Hotel`	0.878 ± 0.022	$\mathbf{0.897} \pm 0.021$	0.828 ± 0.025	0.894 ± 0.012
`Electronics`	0.897 ± 0.041	0.92 ± 0.037	0.912 ± 0.037	$\mathbf{0.944} \pm 0.031$
`Travel`	0.828 ± 0.034	0.848 ± 0.033	0.835 ± 0.025	$\mathbf{0.852} \pm 0.032$
`Software`	0.894 ± 0.025	0.893 ± 0.025	0.919 ± 0.022	$\mathbf{0.956} \pm 0.015$
Pooled AUC	0.876 ± 0.014	0.883 ± 0.013	0.897 ± 0.012	$\mathbf{0.907} \pm 0.012$

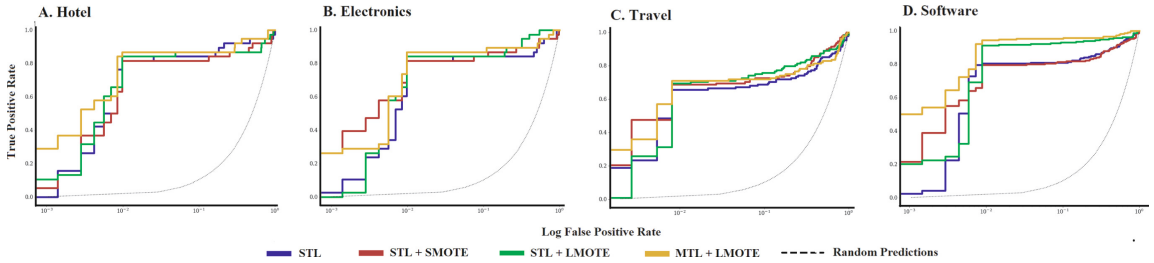


Fig. 3. Domain wise receiver operating characteristics (ROC) curves.

Multi-task Learning Outperforms Single-Task Learning. We compared the performance of our classifier in single and multi-task settings (*STL + LMOTE* and *MTL + LMOTE*) and found that by multi-task learning improves the performance of our classifier. We qualitatively analysed the single and multi task models, and found many instances where by leveraging domain-specific information the multi task model was able to accurately identify suggestions. For instance, consider the following review: *"Bring a Lan cable and charger for your laptop because house-keeping doesn't provide it."* While the review appears to be an assertion (*non-suggestion*), by predicting its domain (`hotel`), the multi-task model was able to accurately classify it as a suggestion.

4 Conclusion

In this work, we proposed a Multi-task learning framework for Open Domain Suggestion Mining along with a novel language model based over-sampling technique for text–LMOTE. Our experiments revealed that Multi-task learning combined with LMOTE over-sampling outperformed considered alternatives in terms of both the F1-score of the suggestion class and AUC.

References

1. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
2. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014). https://doi.org/10.3115/v1/d14-1181
3. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
4. McInnes, L., Healy, J., Melville, J.: UMAP: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
5. Negi, S.: Suggestion mining from text. Ph.D. thesis, National University of Ireland Galway (NUIG) (2019)
6. Negi, S., Daudert, T., Buitelaar, P.: SemEval-2019 task 9: suggestion mining from online reviews and forums. In: SemEval@NAACL-HLT (2019)
7. Peters, M.E., et al.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)

A Novel Fuzzy based Intelligent Demand Side Management for Automated Load Scheduling

Dhairya Karna*, Aditya Vikram, Astitva Kumar, M. Rizwan
Department of Electrical Engineering
Delhi Technological University (DTU)
Delhi, India
E-mail: dhairya_bt2k16@dtu.ac.in

2020 4th International Conference on Green Energy and Applications (ICGEA) 978-1-7281-6021-4/20/$31.00 ©2020 IEEE 10.1109/ICGEA49367.2020.239701

Abstract— **Developments in smart-grid technologies can be associated with rising awareness among general populace of renewable energy as well as the need of distributed generation via these sources. Improvement in efficiency of electrical energy from Renewable Sources (RS) can be achieved by collaborating advanced structures with intelligent metering technology. The smart meters along with the distributed generation sources are being widely used in smart grid applications. An intelligent energy management system is key to monitor and control the processes at consumer and supplier end. Thus, an intelligent system for various computation and procurements can be considered a part of smart-grid. It is within consideration that a part of the energy demand by the building is covered by this Intelligent Demand Management Structure (IDMS). The IDMS is an indispensable tool in order to guarantee greatest added value to the smart meter. The practical and theoretical integration and application of IDMS with the smart meter is presented in this article. The article proposes a novel algorithm based on fuzzy optimization logic employed to the intended system. Fuzzy Controller Logic (FCL) language was used to create the fuzzy rules while the execution was carried out in Python. The designed algorithm is tested in the real time with the load profile of a practical setup. The proposed FCL based algorithm saved a maximum of 15% energy in best cased scenarios.**

Keywords – Smart Meter, IDMS, Fuzzy Optimization, Scheduling.

I. INTRODUCTION

A Lion's Share of residential buildings incorporates distribution and transmission networks as well as low voltage grids connected to independent receivers. For the advancement of electrical energy from renewable sources into residential or industrial buildings there should be an assimilation among the above-mentioned factors. This assimilation would eventually lead to a progressive rise of renewable energy sources for consumers' subscription to energy production. But both these elements bear pros and cons.

One of the pros in expanding the incorporation of renewable sources is the devaluation of non-renewable source consumption while the other is the decrease in the amount of carbon dioxide discharge. Furthermore, it's pragmatic to decrease the energy losses in the grid via perceptive energy transfer management and demand-side optimization in the distinctive prospects of time. A major con of energy contribution via renewable sources in the power sector is the eminent volatility of power generation from such sources. RS discussed in this article specifically includes solar power stations and wind farms. The uncertainty of RS has a strong impact on stabilizing the energy of the power grid and accounts for reduced power quality [1]-[2].

The smart-grid technologies deals with the losses in the grid system as well as use the data to optimize the energy consumption. Regrettably, majority of smart meters that presently employed are simple e-meters with objective of Automated Meter Reading (AMR) and supplementary information procurement. Traditional smart meter discards any data analysis, as it only possesses communication proficiencies and circuit measurement [3]. The primary concern of this system application in such a structure is the need of modernization and integration into further stages for e.g. development of supplier-side Meter Data Management (MDM), which needs concurrent meter replacement in large number of consumer groups [4].

A plausible solution for the above-mentioned dilemma is a perceptive smart-grid that employs a specific codebase on data via an Intelligent Metering System (IMS). It administers real-time data for computation from multiple fields of the power grid. A significant decisive factor of the IMS (aside from Information Technology framework) is the smart meter. Previously, efforts to exploit information from IMS for optimizing energy efficiency and prognosis of demand-side distribution was based solely on data processing via a systematic IT infrastructure. The paper proposes an advanced Intelligent Demand Management Structure (IDMS) for the smart meter technologies with an elective self-governing mode, because of communication losses among several factor affecting the entire system, is beneficial for the safety of the system. Employment of IDMS in smart meters provides an opportunity for inclusion of a steady and multistage modification to the power grid, warranting an equal dissemination of assets in the long term. The aim of undertaking this study was to create a smart meter with a self-governing IDMS. The IDMS used employs a novel fuzzy logic for making rules of the proposed energy management algorithm.

The paper comprises of five different sections. The fuzzy optimization method is briefly explained in Section II of the paper. Section III presents the energy management algorithm proposed for the IDMS. The real time testing results are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:43 UTC from IEEE Xplore. Restrictions apply.

given in Section IV along with the conclusion in Section V of the paper.

II. FUZZY OPTIMIZATION TECHNIQUES

By using appropriate algorithms and data structures, the IDMS permits the optimization of energy efficiency in these buildings. Since a considerable amount of HVAC device/components contribute to the aggregated energy demand in any building, the IDMS might bear tangible gain by saving energy. The IDMS may encompass a plethora of receivers in the structure. The management structure coordinates with the energy source and the storage devices and implements the Demand Response/Demand Side Management structures. It allows for higher efficacy of these mechanisms via automation which would lead to elimination of resident's involvement.

Smart-grid computational techniques include: swarm intelligence, evolutionary algorithms and fuzzy logic [5]. Since results delivered by fuzzy logic is considered to be steady and efficient, we employ it out of the given options [6]. Fuzzy logic is a rationalization of the classic Boolean logic. It was introduced in 1965 with the proposal of fuzzy set theory by L. Zadeh. The fundamental advantage of using fuzzy logic is the ability to interpret the scenarios where no fixed mathematical representation can explain the incident. It allows autonomous decision-making capability in a manner identical to human actions with the assistance from a set of rules that often corresponds to the depiction of the circumstance. Fuzzy Controllers Logic (FCL) and subject blocks involves the following three steps:

Fuzzification: It is the process of transforming the loaded crisp value to a fuzzy value (according to level of membership associated with the fuzzy set). Examples of membership functions shapes are: Gaussian, triangular or trapezoidal. [5].

Inferencing: It is achieved by referring to the fuzzy rule. Formation of the fuzzy rule, due to closeness to the verbal depiction of the incident, permits effortless modelling.

Defuzzification: It is the way with which the precise, crisp values for the output signal is obtained. The result is expressed by the level of activation acquired from the rule database. The most frequent approaches used for sharpening are: Maximum, Center of Area and Centre of Gravity method [7].

The coherent structure of proposed IDMS is shown in Fig1 (Logical Proposed Format for Intelligent Demand Management Structure) has the following measured inputs: a) pv-current; b) batteries-voltage & c) active-power. These input signals are given for fuzzification. Trapezoidal shapes were used for membership functions. Degree of various membership functions are shown in Fig. 2, Fig. 3 and Fig. 4. In the beginning sets of equivalent scopes were used but mid-simulation, they were constantly adjusted. Limitation of these functions were adopted in context of its association with power, daily load, charging characteristics and current characteristics of the photovoltaic power station.

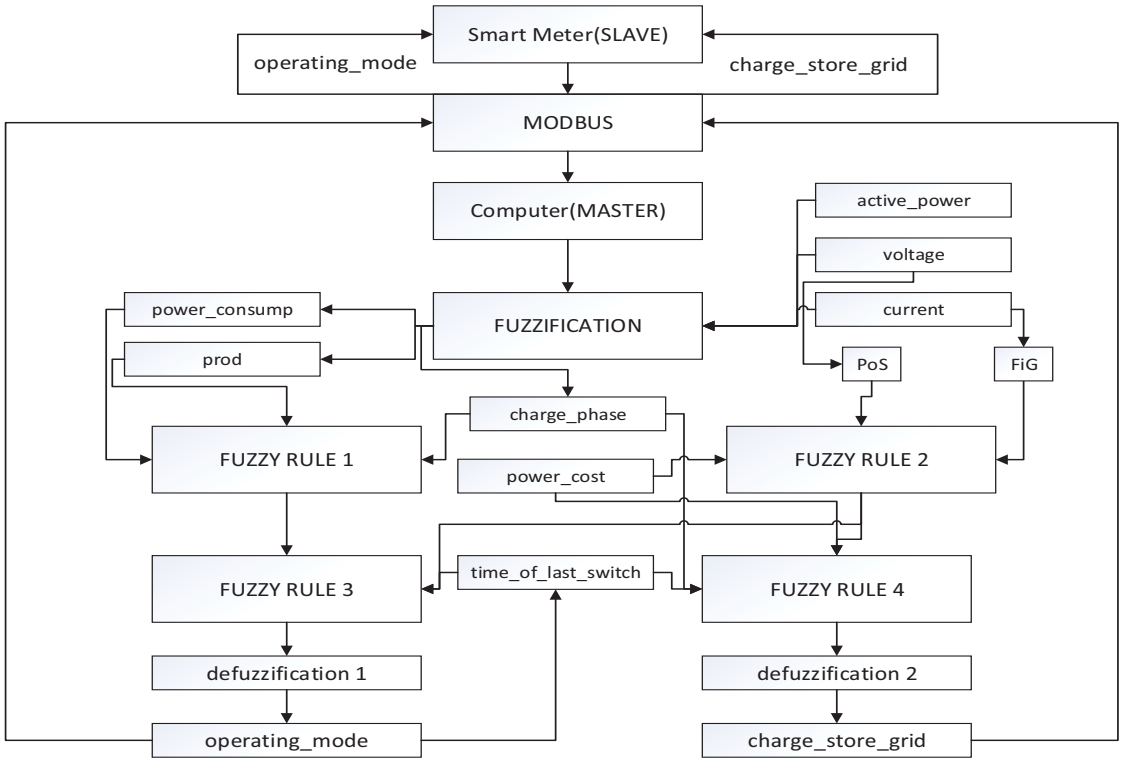


Figure.1. Logical Proposed Format for Intelligent Demand Management Structure

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:43 UTC from IEEE Xplore. Restrictions apply.

TABLE I. PROPOSED RULE TABLE FOR FUZZY NETWORK

FUZZY RULE 1							
IF	Power-consump IS HI	&	Charge-phase IS HI	**&**	Prod IS HI	EQUALS	Later-phase-I IS OFF-NET
IF	Power-consump IS V-LO	&	Charge-phase IS MED	**&**	Prod IS MED	EQUALS	Later-phase-I IS OFF-NET
IF	Power-consump IS MED	&	Charge-phase IS LO	**&**	Prod IS LO	EQUALS	Later-phase-I IS ON-NET
FUZZY RULE 2							
IF	FiG IS NORM	&	PoS IS NORM	**&**	Power-cost IS COSTLY	EQUALS	Later-phase-II IS OFF-NET
IF	FiG IS HI	&	PoS IS NORM	**&**	Power-cost IS COSTLY	EQUALS	Later-phase-II IS OFF-NET
IF	FiG IS LO	&	PoS IS HI	**&**	Power-cost IS COSTLY	EQUALS	Later-phase-II IS ON-NET
FUZZY RULE 3							
IF	Present-phase IS OFF-NET	&	Later-phase-I IS ON-NET	**&**	Later-phase-II IS ON-NET	EQUALS	Present-phase IS ON-NET
IF	Present-phase IS OFF-NET	&	Later-phase-I IS OFF-NET	**&**	Later-phase-II IS ON-NET	EQUALS	Present-phase IS OFF-NET
IF	Present-phase IS ON-NET	&	Later-phase-I IS ON-NET	**&**	Later-phase-II IS OFF-NET	EQUALS	Present-phase IS ON-NET
FUZZY RULE 4							
IF	Charge-phase IS CRITIC	&	Power-cost IS LO COST	**&**	ToLS-batt IS NORM	EQUALS	Bat IS CHARGE-NET
IF	Charge-phase IS CRITIC	&	Power-cost IS NORM	**&**	ToLS-batt IS NORM	EQUALS	Bat IS CHARGE-NET
IF	Charge-phase IS MED	&	Power-cost IS COSTLY	**&**	ToLS-batt IS NORM	EQUALS	Bat IS CHARGE-RS

Post-fuzzification the transformed signals of power-consump, charge-phase, prod was fed to Fuzzy Rule {1}. Due to laborious rectifications and incorporations of the rules the idea of using only one rule database was dropped. In order to start the construction of the rules keeping in mind plausible future rectifications in the structure, a distributed rule database of various blocks was implemented.

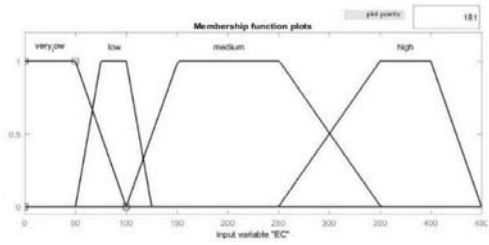


Figure. 2. Energy consumption representation in fuzzy set

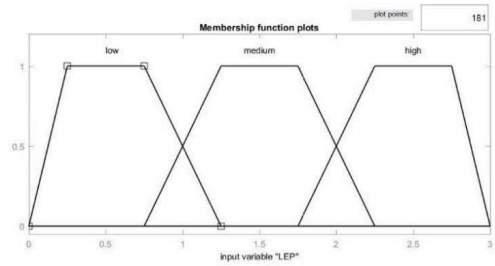


Figure. 3. Phase of charge representation in Fuzzy Set

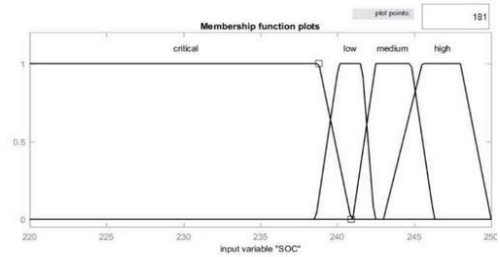


Figure. 4. RS Energy production representation in Fuzzy Set

III. ENERGY MANAGEMENT ALGORITHM

There are four rules as per the proposed algorithm that govern the working of the entire smart meter IDMS, as can be seen from Table 1.

The Fuzzy Rule {1} employed studies the elements correlated to consumption, production and the phase of charge in storage device.

The Fuzzy Rule {2} employed is used for decreasing the amount of times switching occurs in the system. Furthurmore, supplementary variables, associated to the blocks presented in Fig 1(Logical Proposed Format for IDMS): FiG (Fluctuations in Generation), PoS (Price of Storage) and ToLS (Time of Last Switch) are considered. Altogether explanation of their formulas and calculation methods are depicted in Table 2. FiG is calculated for each iteration of 10 min.

The ToLS signal, used in Fuzzy Rule {3}, avoids instantaneous changes in the operating-mode. The power-cost block and the ToLS data is in real-time by the clock of the microcontroller used. Later-phase-1, later-phase-2 and present-phase variables are evaluated in Fuzzy Rule {3}. The output signal from Fuzzy Rule {3} is a Boolean signal that regulates the contactor which is used for controlling the switches used for appliances that requires continuous energy supply. The Defuzzification {1} is associated with the above-mentioned scenario. It's based on comparison that is beneficial for time calculation as the defuzzification operation needs the greatest amount of time and power.

The Fuzzy Rule {4} is employed for regulation of the storage charging process in device. Logged-in data provided to this rule are: ToLS-batt, power-cost and charge-phase. The ToLS-batt variable has its calculation similar to that of ToLS but it controls the contractor that switches the appliances which do not require continuous energy supply. The Defuzzification {2} block is correspondent to the first defuzzification block.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:43 UTC from IEEE Xplore. Restrictions apply.

TABLE II. AUXILIARY SIGNAL CALCULATION

Signal Name	Formula
Price of Storage (PoS)	$PoS = \begin{cases} normal; & SoC > 25\% \\ high; & SoC \leq 25\% \end{cases}$
Fluctuations in Generation (FiG)	$FiG = \begin{cases} low; & V < 25\% \\ normal; & 25\% \leq V < 45\% \\ high; & V \geq 45\% \end{cases}$ $V = \frac{\sigma}{\mu} = \frac{\frac{1}{10}\sum_{k-10}^{k}\sqrt{i_{pv}^{(k)} - i_{pv}^{(avg10min)}}}{\frac{1}{10} \times \sum_{k-10}^{k} i_{pv}^{(k)}} . 100$
Time of Last Switch (ToLS)	$ToLS = \begin{cases} normal; & tolc > 15min \\ short; & tolc \leq 15min \end{cases}$ Tols- time from last change in phase (in minutes) of all the appliances requiring continuous energy supply.
Time of Last Switch Battery (ToLS-bat)	$ToLS_Battery = \begin{cases} normal; & tolc_batt > 15min \\ short; & tolc_batt \leq 15min \end{cases}$ Tols-bat – time from last change of phase (in minutes) of all the appliances which do not require continuous energy supply.

IV. RESULTS AND DISCUSSION

The outcomes from practical tests and simulation of the implemented IDMS which is employed shows an opportunity for enhancement of the daily real-time demand trends. IDMS given in this article, requires a simple framework while not engaging with operating-mode or working time of any regulated device. In best case scenario, the system saved upto 15% (as seen in Fig. 5 & 6). The median for all the observations came out to 7%. It is highlighted that the implemented algorithm in IDMS provided significant cutback in consumption when the demand was at its peak, highlighting its peak-shaving function. Information on Fuzzy Controller Logic (FCL) conventions and its application in Python provides opportunity to integrate and test it further on other environments (e.g. MATLAB).

Since the application is constructed by a common logic and is generalized (as seen in Fig. 7), the algorithm permits larger flexibility and adaptation of the structure to the specified setup. In relation to the given system there is a possibility of power adjustment in different PVs, the varying capacities of storages used or the demand of power. This structure is also beneficial for the whole power system.

The IDMS offers the advantage of choosing a range of storage devices with lower capacities as well as PVs with lower output power in comparison to the standards renewable energy source with storage solutions. The algorithm for energy management allows the collaboration of storage element and the local source of energy which is launched in the smart meter. After that it was tested using real-time data.

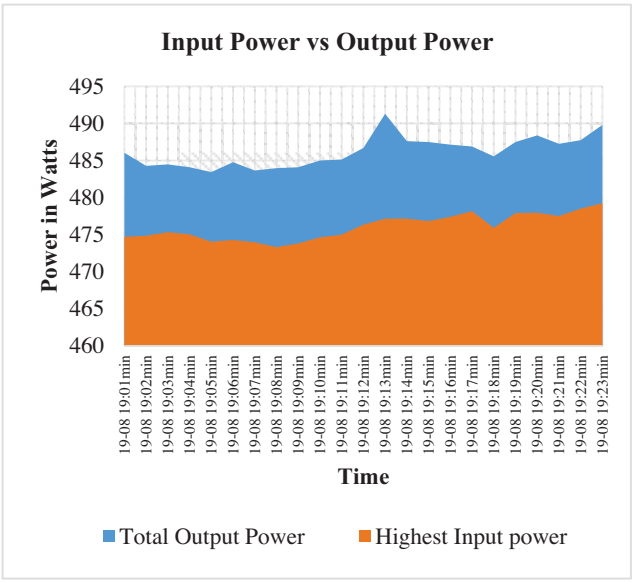


Figure. 5. Graph representing value of highest input (in that minute) vs total output power after IDMS implementation

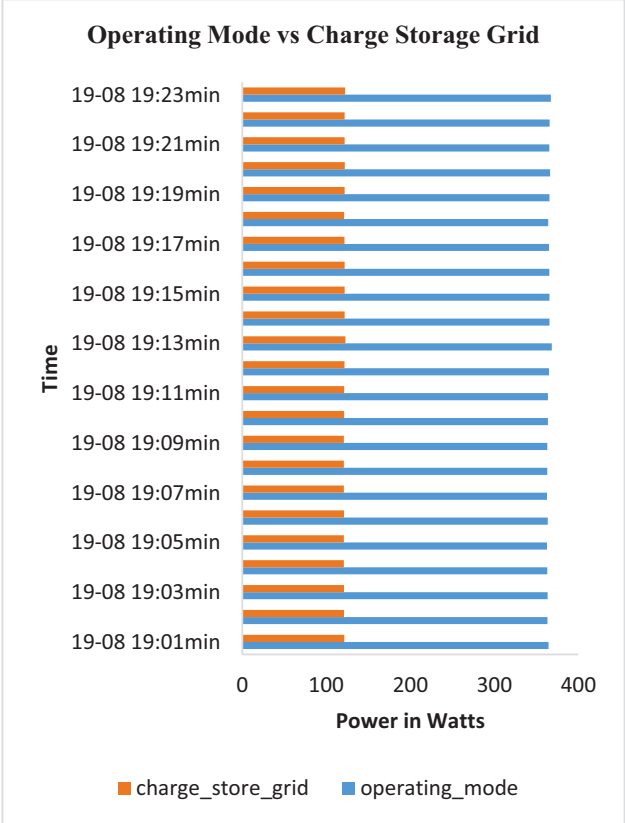


Figure. 6. Graph representing operating mode and charge store grid values (two outputs after defuzzification of input variables.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:43 UTC from IEEE Xplore. Restrictions apply.



Figure. 6. Smart Energy Meter

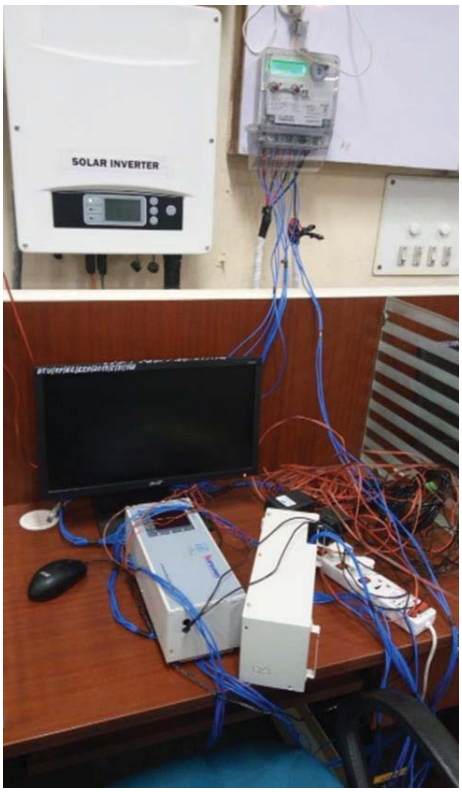


Figure. 7. Data monitoring system integrated to smart meter

V. CONCLUSION

In this paper, the traditional smart meter was provided with computational intelligence by implementing an IDMS which uses fuzzy based algorithms. The results show the peak-shaving properties provided by the smart meter and intelligent use of electrical energy. Financially this arrangement offers the advantage to the consumer of peak shaving and regulated price of storage by using RS. Furthermore, the decreased value of maximum power also indicates that the transmission losses are minimized. Further integration with an Android Application is a feasible and reasonable argument so as to decrease the cost of power consumption as well as provision of remote switching capabilities.

REFERENCES

[1] Y. Yan, Y. Qian, H. Sharif, D. Tipper, A Survey on Smart Grid CommunicationInfrastructures: Motivations, Requirements and Challenges, IEEE Communications Surveys & Tutorials, 15, 2013, pp. 5–20

[2] A. Kumar, M. Rizwan, U. Nangia, A Hybrid Intelligent Approach for Solar Photovoltaic Power Forecasting: Impact of Aerosol Data, Arabian Journal for Science and Engineering, 2019, pp. 1-18

[3] A. Ghelli, H. Hagras, G. Aldabbagh, A Fuzzy Logic-Based Retrofit System for Enabling Smart Energy-Efficient Electric Cookers, IEEE Transactions on Fuzzy Systems, 23, 2015, pp. 1984–1997

[4] A. Chojeki, M. Rodak, A. Ambroziak, P.Borkowski, Energy Management System for Residential Buildings, 90-537 Lodz, Poland, IET, IET Research Journals, 2015, pp. 1-13

[5] P. Rajendhar, B.E. Jeyaraj, Application of DR and co-simulation approach for renewable integrated HEMS: a review, IET Geneartion, Transmission Distribution, 2019

[6] A. F. Zobaa, A. Vaccaro, Computational Intelligence Application in Smart Grids : Enabling Methodologies for Proactive and Self Organizing Power Systems, London, Hackensack, NJ: Imperial College Press, 2015

[7] M. Rohmanuddin, Thee Houw-Liong, Ahmad, Y.Y. Nazaruddin, Simplifying fuzzy rule base of multiple input multiple output systems by construction multi-layer fuzzy controller, IEEE, 2000, pp. 3728–3733.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:53:43 UTC from IEEE Xplore. Restrictions apply.

Cluster Computing
https://doi.org/10.1007/s10586-020-03115-0

A route selection approach for variable data transmission in wireless sensor networks

Aarti Jain[1] · Manju Khari[1] · Elena Verdú[2] · Shigeru Omatsu[3] · Rubén González Crespo[2]

Received: 25 February 2020 / Revised: 19 March 2020 / Accepted: 17 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
The nodes in wireless sensor networks (WSNs) are responsible for communicating data which is primarily of three types viz. video, audio and text. In literature, a large number of energy aware and shortest path based route selection approaches have been proposed to enhance the routing efficiency in WSNs. Most of these route selection approaches are designed by assuming fixed type and size of data packets and results in the same path selection for all types of data. This path selection, does not consider data size and data priority, results in high contention losses, topology failures and non-uniformity in energy depletion of nodes. In this paper, a route selection approach which is based on the preservation of network connectivity for improving overall network lifetime has been proposed for applications where sensor nodes are required to transmit different types of data. The proposed approach preserve highly connected edges at the initial rounds of data communication in an energy efficient manner, such that network connectivity would be maintained even at later rounds of data communications. The proposed route selection approach, VDR, (Variable Data rate Routing) has been simulated with both proactive and reactive route maintenance strategies. The simulation results show that the proposed algorithm with both proactive and reactive route maintenance strategies result in better network lifetime, energy consumption, overall network connectivity, packet delivery ratio as compared to the existing State-of-the-Art energy efficient route selection approaches with proactive and reactive route maintenance strategies. The results also show that the proposed algorithm is scalable and performs better than the compared algorithms both at low node density and high node density.

Keywords Wireless sensor networks · Routing optimization · Energy efficiency · Connectivity measure · Variable data rate · Network lifetime

1 Introduction

Rapid increase in wireless communication technology and growth in production of low cost & small wireless communication devices have driven the researchers to explore more protocols in the field of WSNs [1]. The main responsibility of the nodes in WSNs is to collect information from the target area and then transmit the collected data to the intended user via sink node. In WSNs, multi-hop routing of data (from source node to sink node) is preferred instead of one hop communication method [2] that encourages the researchers to design and improve the multi-hop routing methods which are more energy efficient, connectivity aware and ascertain the required QoS.

In WSNs, collected information is primarily of three types viz. video, audio and text. Different types of data viz. video, audio and text vary in their respective data size and priority of transmission. Moreover, as WSNs are mostly deployed for monitoring applications, the communication of audio and video messages is delay bound rather than delay sensitive. In WSNs, data in the form of text is of small size and usually used for controlling and setting up of routing paths, whereas audio and video messages are of

✉ Rubén González Crespo
ruben.gonzalez@unir.net

[1] Ambedkar Institute of Advanced Communication Technologies and Research, Geeta Colony, New Delhi, Delhi 110031, India

[2] Faculty of Engineering, Universidad Internacional de La Rioja, Av. de la Paz, 137, 26006 Logroño, La Rioja, Spain

[3] Digital Manufacturing Education and Research Center, Hiroshima University, 3-10-31 kagamiyama Higashi-Hiroshima, Hirodhims 739-0046, Japan

Published online: 27 April 2020

large sizes and are required to be transmitted to sink node while giving more emphasis to energy efficiency and reliability. For these different forms of information, consumption of energy for the transmission of data also varies.

It has been observed that in most of the energy efficient multi-hop routing protocols are based on shortest path algorithm. The selection of next hop node is made irrespective of the data size; which is low for text transmission and high for video and image transmissions. Also, the selection of next hop node is irrespective of the role of the next hop node in maintaining network connectivity. Consequently, for a large set of data transmission for video and image transmission, some of the most connected nodes which are part of many shortest paths from sensor nodes to sink node(s), deplete their energy much before than other less connected nods and thus leads to failure of these nodes. This early failure leads to creation of energy holes as these highly connected nodes disconnect some of the nodes from rest of the network and thus coverage and connectivity of a significant part of the network is vanished and the network lifetime terminates prematurely. Moreover, if most of the sensor nodes use these highly connected nodes present on the shortest path to route the data, it will lead to contention and congestion losses. Hence for designing of efficient[1] routing protocol for WSNs, it is important to consider the various types of data to be transmitted and role of next hop nodes in maintain connectivity of the network.

To address these issues "A Route Selection Approach for Variable Data Transmission for Wireless sensor Networks" has been proposed. In the proposed method, while the Text packets/small set of data (control and short messages) are transmitted by shortest path, for video and audio data transmission (large set of data) connectivity metric of edges/nodes is considered for route selection. This method is applicable for video and audio transmission in WSNs as in these networks, unlike, wireless networks, transmission of video and image messages is delay bound rather than delay sensitive and Text messages are mainly used for carrying out route maintenance and updating. Hence accordingly, in the proposed method, high size data (video and Audio) is usually transmitted by minimum connected links/nodes. By applying this methodology, highly connected edges/nodes are preserved at the initial stage to maintain connectivity at later stages of the network lifetime and network remains connected for longer time [3, 4] vis-a-vis. network lifetime get enhanced.

In the proposed method, every node in the network has been associated with a degree of connectivity. This degree of connectivity has been derived by using proposed centrality index. Further based on the proposed centrality index, nodes are classified into three categories i.e. are high, medium and low and are represented as θ, Ω and η nodes respectively. Hence, η nodes are highly connected, Ω are moderately connected and θ are low connected nodes. This connectivity degree of a link/node is used while route selection so that network connectivity can be maintained for longer duration. The proposed association of degree of connectivity with nodes does not led to increase in computational complexity of route selection method and efficiently satisfy the proposed need of preserving highly connected nodes at the initial stage of the network lifetime to enhance the connectivity viz. a viz. the effective network lifetime.

The main contributions of this paper are as follows:

- Modelling of WSN as weighted graph. Our modelling approach considers the geodesic distance between respective nodes to define weights of the network edges.
- Deriving the parameters for network connectivity and to address the constraints related to network connectivity.
- Proposed a stand-alone distributed route selection algorithm that considers both minimization of cost of communications and maximization of time to network partition.
- The proposed routing method is developed for efficient path selection for variable data rate transmission and can be associated with reactive or proactive route maintenance strategies.

The proposed route selection algorithm i.e. VDR is simulated using MATLAB and performance evaluation has been done with both reactive and proactive route maintenance strategies. The simulation results demonstrate that the proposed algorithm resulted in prolonged network lifetime, connectivity ratio, end-to-end delay, packet delivery ratio and load distribution in comparison to TSRA; a Tabu search based energy efficient reactive routing protocol and EERP.

The paper is organized as follows: Sect. 2 presents the related work. Section 3 presents the system model, methodology to categorize the different links in different types of connectivity arcs and details of the proposed algorithm. Simulation studies and performance analysis are presented in Sect. 4. The concluding remarks are presented in Sect. 5.

2 Related work

In WSNs, the most common aim of the proposed routing protocols is to find out a routing path that can route data reliably to the sink while ensuring improved network lifetime. In order to improve the network lifetime several

[1] Efficient routing is defined as the routing methodology which is energy efficient, having high throughput and provide QoS.

authors have developed energy-efficient algorithms (both reactive and proactive) for multi-hop communications in WSNs [5–12]. The overall objective of these algorithms is to either maximize the first node die (FND) lifetime metric or to maximize the overall network liftime. EAODV [13] routing algorithm is an energy efficient version of AODV [14] routing protocol. In EAODV, the source to destination path with lesser energy consumption is selected for data transmission. The performance of EAODV is better than the AODV in terms of network lifetime. However, the delay associated with setting up of a path is high, as algorithm has used IDSA (Iterative dimidiate search arithmetic) for selecting proper transmission power. Heo *et. al.* have proposed EARQ [15] which is a proactive energy efficient routing protocol. It is an energy efficient routing protocol for real-time, reliable delivery of a packet. In EARQ, a node proactively collects information related to energy cost, delay and reliability of a path to destination. Based on the collected information, path to destination is selected.

Singh et al. [16] have proposed five metrics for the selection of the routing path for energy efficient routing. These metrics are minimum-energy, maximization of time to network partition, minimization of variance in node energy levels, minimize of transmission cost and minimization of maximum node cost. Out of the five metrics, only the first (minimum-energy path) and fourth (minimize node cost) have been implemented. Park et al. [17]. considers the MMBCR (min–max battery cost routing) and CMMBCR (conditional MMBCR) online algorithms to select a source-to-destination path. The MMBCR algorithm selects a path P for which the minimum of the residual energies (i.e., energy remaining following a route) of the sensors on path P is maximum. In CMMBCR, a source-to-destination path is selected for in which no sensor node on that path has residual energy below a threshold γ. If there is no source-to destination path with this property, then the MMBCR path is used. In all of these algorithms Dijkstra's shortest path algorithm is used to select the paths. Although these metrics tends to minimize the total (or average) energy consumed over a sequence of routes, they do not focus on the primary objective of maximizing lifetime. Orojloo and Haghighat [18] have proposed a reactive Tabu search algorithm based heuristic routing algorithm (TSRA) for WSNs. The main objective of this algorithm is to minimize the energy consumption. In the proposed algorithm the network is modelled as a weighted graph. The weight assigned to the edges of the graph is on the basis of the node's residual energy and the transmission cost. Finally on the basis of weight of edges shortest path from source nodes to sink node are determined.

In [19], Chang and Tassiulas *et. al.* have proposed a shortest cost path based routing algorithm. The algorithm has considered link cost in terms of both the communication energy consumption rates and the residual energy levels for maximizing network lifetime. Akkaya et al. [20] and [21] have proposed an energy efficient routing protocol which selects a low-cost and delay-constrained path for real-time packets in WSNs. The protocol assumes that every node knows the position of all other nodes and also the cost of transmission. These algorithms have mainly assigned weights to the links on the basis of communication cost, energy or delay and have applied Dijkstra's shortest path algorithm to find the path. Mohajerani et al. [22] have proposed ACO based routing algorithm (LTAWSN for WSNs. In this routing algorithm, the pheromone accumulation value is made dependent upon the number of hops and energy consumption by nodes falling on a route from source to sink node.

In literature several multipath routing protocols have also been proposed that allow each source node to find multiple paths towards the sink node. The state-of-the-art multipath routing protocols are Braided Multipath Routing [23], N-to-1 Multipath Routing Protocol [24] and EEMRP [25]. In these multipath based routing protocols, multiple node-disjoint paths and a load balancing method is used to assign the traffic over each selected path. Boulfekhar and Benmohammed [26] have proposed a proactive, energy aware multipath routing algorithm (EERP) for WSNs. In the proposed algorithm, firstly controlled flooding of beacon packets is executed by nodes to find the local topology, energy cost and all possible routes between nodes and sink node. Finally, multiple routes are maintained between nodes and sink by using periodic flooding of beacon messages.

In literature several cluster based routing protocols have been proposed to enhance the network lifetime. In [27], Heinzelman *et.al.* have proposed a Cluster-based and self-organizing dynamic routing protocol LEACH for WSNs. The protocol has the ability to perform re-clustering. However, the concept of dynamic and re-clustering results in extra overhead. In literature many variants of LEACH protocol are proposed like BARC [28], AZR-LEACH [29], EHE-LEACH [30], TEEN [31], APTEEN [32], PEGASIS [33], and TTDD [34], for energy efficient routing. However, it has been observed that LEACH based routing protocols results in unbalanced energy consumption among the nodes and which results in premature end of network lifetime.

In [35], a joint global and local route update method is proposed for proactively updating and maintaining the routes for energy efficient communications in clustered

WSN. In the proposed method, PRR (Packets Receiving Rate) has been used for global route maintenance between Cluster heads and the sink node, while local update process uses the link quality metric to keep check on the route failures. The frequency of executing the local route update is made higher than global route update. However, in case, the local route update method founds substantial changes in routes, it informs the global route method for early route performance evaluation and maintenance.

In [36], a new cluster head election protocol SEP has been proposed to prolong the network lifetime without any node failure, which has been termed as stability period. The selection of cluster head is made dependant on the residual energy. The simulation results show that the proposed method resulted in enhanced stability period. However, the increase in stability period results from the deployment of heterogeneous nodes in the network i.e. deployment of few nodes with more initial energy level than other deployed nodes.

Subsequently AlanaziA et al. [37] have proposed a modified SEP protocol in which again heterogeneous node deployment is considered, however node degree has considered as an additional cluster head selection parameter. P-SEP [38] is another modified SEP protocol for enhancing the stability period of the network. In P-SEP, the cluster head election is similar to [37], but it has used fog infrastructure to provide location awareness and link quality between nodes.

From literature survey, It can be observed that prolongation of network lifetime is achieved by selecting data transmission routes on the basis of residual energy, transmission cost, path length; multipath routing; clustering of nodes; combination of these methods. It can also observe that in [16] Singh et al. have proposed five metrics viz. minimum-energy, maximization of time to network partition, minimization of variance in node energy levels, minimize of transmission cost and minimization of maximum node cost for the selection of the routing path for energy efficient routing. Out of the five metrics, only the first (minimum-energy path) and fourth (minimize node cost) have been implemented. The proposed work in this paper focuses upon maximization of time to network partition to enhance the network lifetime. The prolongation of time to network partition has been achieved by preserving the overall network connectivity. In the proposed method, the future network connectivity has been considered while selecting the energy efficient and shortest paths for different types of data. By maximizing the time to network partition for the whole network lifetime, overall performance of the network in terms of network lifetime, packet delivery ratio has been considerably improved.

3 System model and proposed algorithm

This section is consist of three subsections. The first subsection presents the WSN system model and other network related considerations. Second sub-section presents the categorization of edges into different connectivity arcs and third subsection presents the proposed connectivity based routing algorithm.

3.1 System model

In the proposed method, the WSN has been modeled as edge weighted randomly distributed Graph *G(V,E)*. Here *V* represents the set of sensor nodes and *E* represents set of edges. For each edge *(u,v)*, a function is defined to map the distance between nodes to its weight value $w(u,v)$. Thus, Each Edge e in set E is associated weight w$\in$ W. The definition of weight of edge in randomly distributed graph is given as:

Definition 1 (*weight of an edge*) Weight of an edge (W) is a set of ordered pairs such that for any edge $\{(\mathrm{u},\mathrm{v}),\mathrm{w}(\mathrm{u},\mathrm{v}))|(\mathrm{u},\mathrm{v}) \in \mathrm{E}\}$, the $w(u,v)$ is defined by following equations:

$$w(u,v) = \begin{cases} 0 & for \quad d_{u,v} > r^c_{\max} \\ (r^c_{\max} - d_{u,v})/(r^c_{\max} - r_{cr}) & for \quad r_{cr} \le d_{u,v} \le r^c_{\max} \\ 1 & for \quad d_{u,v} \le r_{cr} \end{cases} \quad \forall (u,v) \in E \tag{1}$$

Here, $d_{u,v}$ represents the actual distance between node u and node v. $r^c_{\max}$ represents the maximum permissible communication radius of a node. r_{cr} represents the minimum distance between node u and node v depending upon node density and network area.

According to definition 1, every edge (u,v) a weight value $w(u,v) \rightarrow [0,1][0,1]$ is assigned, where $w(u,v) = 0$ represents edge (u,v) does not belongs to G and $w(u,v) = 1$ represents the edge (u,v) fully belongs to graph G. All other cases represent real edges. A high value of $w(u,v)$ i.e. near to 1 represents a smaller distance between nodes u and v.

For radio transmission and reception, we have considered first order radio propagation model [39]. According to the model, transmitting and receiving costs are computed as follows:

$$E_{Tx}(b,d) = \begin{cases} b * E_{elec} + b * \varepsilon_{fs} * d^2, d < d_0 \\ b * E_{elec} + b * \varepsilon_{amp} * d^4, d \ge d_0 \end{cases} \tag{2}$$

$$E_{Rx}(l) = l * E^{Rx}_{elec} \tag{3}$$

Here, d_0 represents the fraunhofer distance, d is the actual distance from sender node to receiver node. b represents,

number of bits per signal, E_{elec} is the circuit's energy dissipation. ε_{fs} and ε_{amp} are the energy consumption factor for the free space and multipath amplification respectively. After initialization of network, each node $u \in V$ identifies its respective neighbor nodes N_u. N_u are those nodes for which $w(u,v) : V \times V \geq 0$.

3.2 Categorization of edges into connectivity arcs

The importance of edge between two nodes and its role in network connectivity plays an important role in both theory and applications of randomly distributed graphs [40, 41]. In [4], the different edges between nodes has been categorized into three categories by using Betweenness Centrality for maximizing the network lifetime. However, the computational complexity for calculating Betweenness centrality is too high to be executed for nodes with low computing capacity in WSNs. In this method, like [4] we have also categorized the nodes into three categories, however the proposed categorization has used node degree and distance from sink node to define the connectivity strength of a node. The details the three types of connectivity links namely θ, Ω and η arcs as defined below:

Definition 2 (*connectivity strength of an edge*) For a given random graph G, the connectivity strength of any edge connecting node x and node y, is calculated as:

$$CONN_G(x,y) = \frac{N_x^n + N_y^n}{d_{x-s}^n + d_{y-s}^n} \tag{4}$$

where d_{x-s}^n and d_{y-s}^n represent the normalized distance between nodes x and y from the nearest sink node (destination node) respectively. N_x^n and N_y^n represent the normalized number of neighbour nodes of node x and y respectively.

The connectivity strength proposed above has considered the (1) number of nodes which are connected to each other via using edge (u,v) between nodes and (2) distance of nodes from the sink node. The node degree of nodes is directly related to connectivity strength as more the number of nodes connected through an edge more is the connectivity strength. The distance from the sink node is important as the node closer to sink nodes act as relay nodes and hence are more important for maintaining connectivity.

To categorize the different edges into different types of connectivity arcs, firstly the edge between sink node and its neighbour node, having maximum connectivity strength $CONN_G(x, \sin k)$ is categorised as η arc (highly connected arc) and the connectivity strength for the same has been assigned to a variable termed as $CONN_G(\eta)$. The edge between sink node and its neighbour node, having minimum connectivity strength $CONN_G(x, \sin k)$ is categorised as Ω arc (moderately connected arc) and the connectivity strength for the same has been assigned to a variable termed as $CONN_G(\Omega)$.

The categorization the other edges have been done as follows:

- An edge (u, v) in the graph is categorized as θ arc if
$$CONN_G(u,v) < CONN_G(\Omega) \tag{5}$$
The edges which are classified as θ arcs have minimum network connectivity. These types of edges have minimal effect on the network connectivity.
- An edge (u,v) in the graph is called η arc if
$$CONN(u,v) \geq CONN_G(\eta) \tag{6}$$

Such types of arcs are highly connected arcs. Early disconnection of these arcs may lead to network partition and which in turn hampers the overall network performance.

Next section presents the proposed route finding algorithm which have used the connectivity arcs θ, Ω and η, data type and multiple sinks to find an optimal path for final communication. The goal of the proposed routing algorithm is to efficiently route each routing request r_h, consisting of different types of data viz. text, audio and video, in such a manner that it maximizes the number of successful routing requests.

3.3 Proposed algorithm

The proposed algorithm optimizes the energy consumption and maximizes network lifetime for variable data transmission by using degree of connectedness and Dijkstra's shortest path algorithm. In the proposed algorithm, for the given network modeled as graph *G(V,E)*, after initial deployment, on the basis of the degree of connectedness (Sect. 3), all the edges are categorized into θ, Ω and η arcs. Where η arcs are highly connected edges, Ω are moderately connected and θ edges have low connectivity.

At the time of initiation of routing request $r_h(x,y)$, first of all, the type of data to be transmitted is identified. For video and audio data type, the total edge weight and the degree of connectedness is considered for shortest path selection. However, for text data, only edge weight based shortest path is considered. The same has been done as text only data has lower data set as compared to audio and video data type and has lesser effect on energy parameters of a node. Moreover, in WSNs the communication of audio and video messages is delay bound rather than delay sensitive.

To route video and audio data, all possible shortest path k are identified to sink node. Among all the k paths, the

final path selection is made on the basis of θ, Ω and η connectivity arcs. The highest preference is given to paths having least number of highly connected edges i.e. η edges. Among the paths having same number of highly connected edges, the preference is given to routes with less number of Ω edges. The objective of considering connectivity of edges is to preserve the highly and moderately connected edges so that network connectivity can be prolonged. The pseudo code for the proposed algorithm is as follows:

Proposed Algorithm (VDR)

```
1.  ∀ nodes V∈ G(V,E)
2.        assign σ energy to all nodes ;
3.  ∀ edges E_i(u,v) ∈ G(V,E)
4.  ∀ edges Categorize into θ, Ω and η arc
5.  For each node V_i ∈ G(V,E)
6.        if (data_ready = transmit)
7.            source ← x ←V_i;
8.            destination ← y;
9.            assign packet_size;
10.       elseif
11.           wait and return to step 5;
12.       endif
13. for x
14.   if (packet_size ,N = 48)
15.       data_type=text
16.       Dijkstra's(x,y,N,w_t)
17.   elseif (Packet_size = 128 || Packet_size = 256)
18.       data_type=audio || video
19.           Dijkstra's(x,y,N,w)
20.               select optimal path among k paths on the basis of θ, Ω and η arcs, avoiding η arcs ;
                      consider the path having minimum connectivity;
                          transmit data packet
21.   endif
22. calculate residual energy
23. calculate network lifetime
24. calculate energy consumed
25. return line 5
```

For the selection of path, Dijkstra's shortest path algorithm has been used and considering the degree of connectedness as another route selection parameter leads to optimization of both network lifetime and energy consumption.

3.3.1 Details of the proposed algorithm

In the proposed algorithm VDR, line 1 to 4 represents the modeling of the network as edge weighted randomly distributed graph and categorization of edges into different connectivity arcs. Line 5–9, represents the generation of data by a node. Lines 14–19, represent the data size identification viz. text, image or video and according to data size, the path to destination is selected (line 20). Lines (22–25) represent the updating of network parameters for future transmissions. Due to consideration of connectivity strength (preferring edges with low connectivity) to find the route between nodes and sink node has resulted in uniform load distribution among the routes available for routing between nodes. This consideration of connectivity

strength hence resulted in less congestion and contention losses along with network longevity.

4 Performance evaluation

The proposed routing method with both reactive and proactive route discovery has been simulated by using MATLAB to evaluate their performance. The performance of the proposed method has been compared with TSRA [18] and EERP [26] which represent the recently proposed energy efficient reactive and proactive routing protocols respectively. TSRA is a Tabu search algorithm based extension of reactive routing protocol AODV [14]. TSRA considers the intergration of energy consumption and number of hops while selecting a path. In order to avoid the selection of forwarding nodes with less energy, the proposed method forms Tabu-list and Tabu-tenure and use them for optimized path selection. EERP is an energy efficient proactive routing protocol developed for wireless sensor networks. In EERP, In EERP, each sensor node maintain a neighbour table and a list of suitable neighbours through which it can transmit packet to base station. The selection of neighbour for data forwarding is done on the basis of minimum cost function and active state function.

For the simulation analysis, wireless sensor network is modeled as a randomly distributed graph. The simulation parameters and implementation methodology are discussed in Sect. 5.1. In Sect. 5.2, the performance of the proposed method has been compared with TSRA and EERP in terms of network lifetime, packet delivery ratio, first node die (FND), residual energy distribution and end-to-end delay.

4.1 Simulation environment

For simulations, a wireless sensor network consisting of N sensor nodes and M base station has been considered. The N sensor nodes are deployed randomly and base station is located randomly in a field of size 1000×1000 m^2. All the nodes have been initiated with initial energy σ equals to 1 J. Nodes in the network randomly generate multiple messages of different types (i.e. text, audio or video). The control and text data packet sizes are fixed and kept equal to 48 bytes, audio packet size is kept as 128 byte and video packet size is kept as 256 bytes. For simulations the Log distance path loss energy model has been used [39].

To evaluate the performance, 50 different topologies have been generated. The performance of all the compared methods has been evaluated 100 times for each topology. The performance is evaluated for simulation consisting of 500 runs of communication period. Finally, average value of the performance metrics has been considered for performance evaluation and comparison. For proactive route maintenance, hello messages are broadcasted after every 10 s of the simulation time. In addition to these, Table 1 summarizes the other important simulation parameters considered for simulations.

4.2 Results

In this section, we have analyzed the simulation results. The evaluation of the performance of the four methods is done in terms of network life time, end-to-end delay, connectivity percentage, packet delivery ratio and load distribution.

4.2.1 Network lifetime

Network lifetime i.e. number of alive nodes versus simulation time is an important metric to check the efficiency of any method. The performance is evaluated at different node densities. For simulations, the minimum and maximum transmission ranges of nodes are set to 30 m and 150 m. The same is in accordance with the optimal communication range for MICAZ motes [39, 41] respectively. Figure 1a and b presents resulted network lifetime for all methods for node density 100 and 200 respectively.

From the simulation results, it has been observed that for both values of node densities; the proposed method with reactive route maintenance R-VDR results in highest network lifetime as compared to proposed method with proactive route maintenance P-VDR, TSRA and EERP. It can also be observed that the network lifetime for P-VDR is better than EERP. The enhanced performance of the proposed methods can be attributed to even load distribution among the nodes and lower contention losses.

Moreover, it can be observed from Fig. 1a and b that the network lifetime of R-VDR and P-VDR has improved at higher node density. The improved performance at high value of r_t can be attributed to higher network connectivity with increase in node density which leads to more optimal path selection.

4.2.2 Connectivity percentage

This metric examines the percentage of alive nodes having connectivity up to sink node and indicates the efficiency of the protocol in maintaining uniform energy consumption

Table 1 Simulation parameters

E_{elec}	100 nJ/bit
E_{amp}	50 pJ/bit/m^4
Route timeout	10 s
Minimum required energy(E_{min})	50 *nJ*

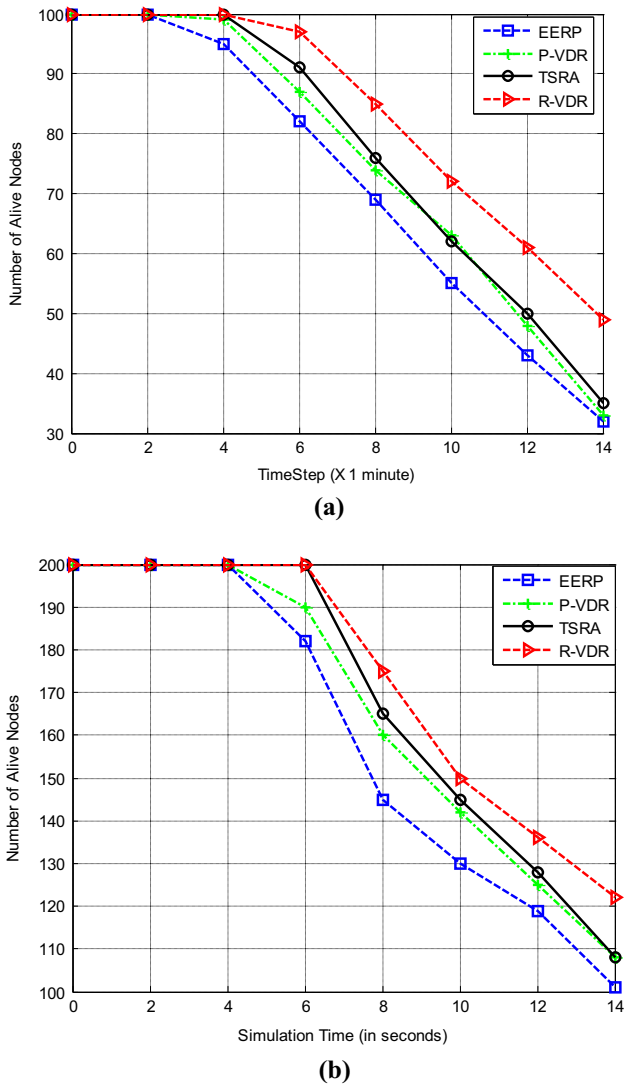


Fig. 1 Lifetime comparison between R-VDR, P-VDR, TSRA and EERP routing algorithms at node densities **a** N = 100 **b** N = 200

and load distribution. It has been observed that though there are certain nodes which have sufficient energy to communicate with neighbor nodes, due to non-availability of intermediate nodes these nodes are not able to communicate with the sink nodes. Figure 2a and b shows the number of percentage these isolated nodes in terms of remaining nodes in the network for node density 100 and 200 respectively. It can be observed that R-VDR and P-VDR results in lower number of isolated nodes and thus results in better performance of the network.

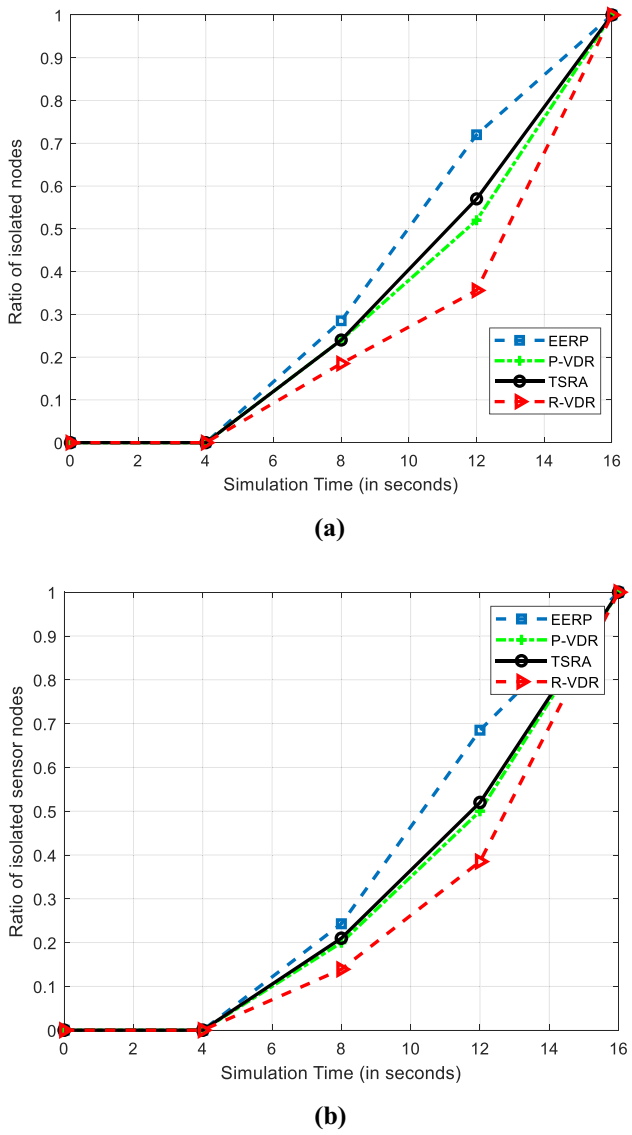


Fig. 2 Percentage of alive nodes with no connectivity to base station. **a** N = 100, **b** N = 200

4.2.3 Packet delivery ratio

In Fig. 3a and b the performance of R-VDR, TSRA, P-VDR and EERP has been analyzed in terms of packet delivery ratio in respect to simulation time. Packet delivery ratio metric examines the efficiency of a routing method in terms of number of successfully delivered packets to the sink. As the simulation time advances, the energy of nodes in the network start depleting which results in failure of

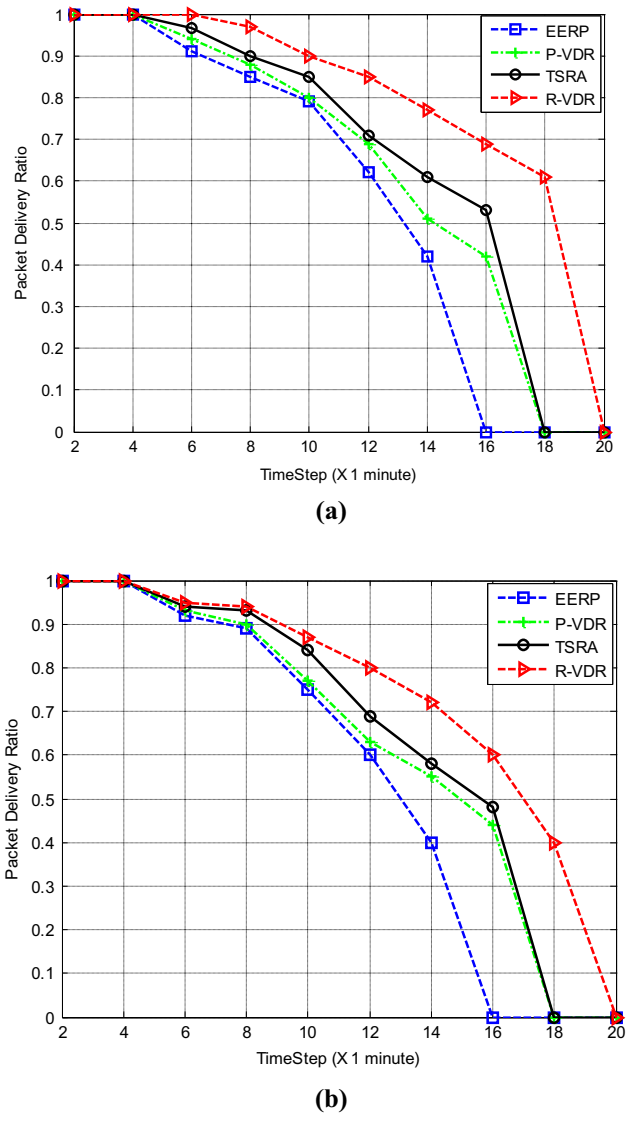


Fig. 3 Successful packet transmission with respect to simulation time. **a** N = 100, **b** N = 200

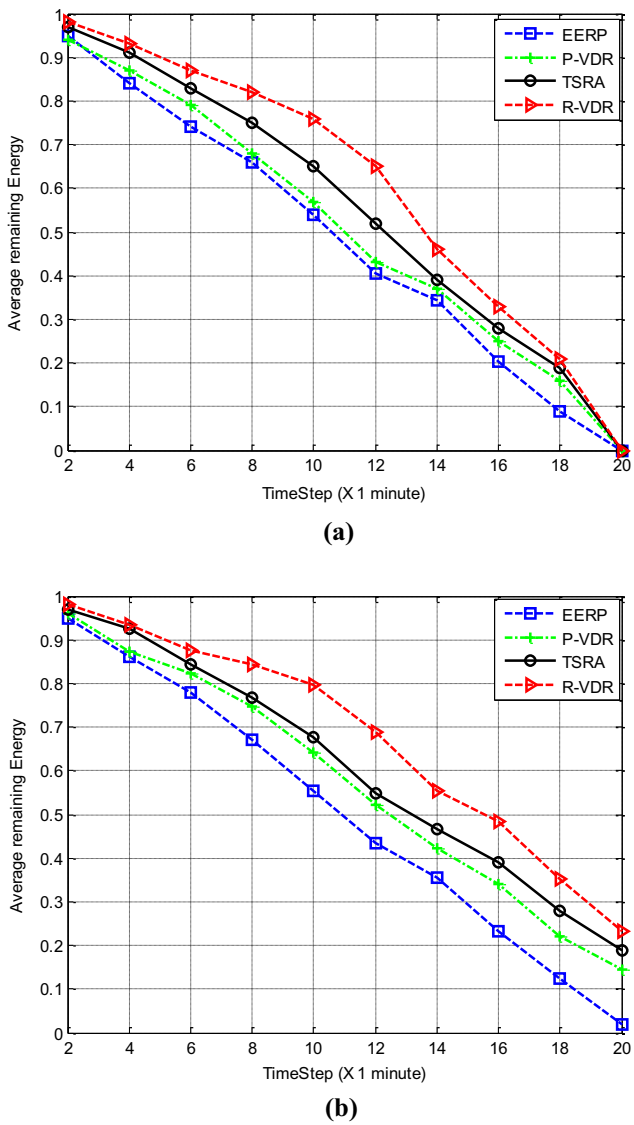


Fig. 4 Comparison of average remaining energy after every 2 min of simulation time for **a** N = 100, **b** N = 200

certain nodes. This failure of some nodes affects the network connectivity, which in turn adversely impacts the number of successful receptions. It can be observed from the figure, that overall packet delivery ratio (video, audio and text) of the proposed P-VDR and R-VDR is better than TSRA and EERP respectively. The better performance in terms of packet delivery ratio is due to higher network connectivity in the respective subsequent rounds.

4.2.4 Load distribution

Load distribution characteristics are observed on the basis of average remaining energy and standard deviation in the remaining energy of the deployed nodes after every 2 min of the simulated time. Figure 4a and b shows the average remaining energy of the sensor nodes with respect to simulated time. The simulation results indicate that with increase in simulation time, the average remaining energy of R-VDR and P-VDR is comparable with TSRA and EERP respectively. However, Table 2 indicates that in case of both R-VDR and P-VDR, the value of standard deviation in remaining energy of nodes is lower than EERP and TSRA. This validates that the proposed methods result in fair load distribution and preservation of highly connected nodes for future communications.

4.2.5 End to end transmission delay

Table 2 shows the standard deviation in remaining energy of nodes at the end of every 2 min of simulation time. In the proposed method, the communication load is diverted

Table 2 Standard deviation in remaining energy versus simulation time

Simulation time (X 1 min)	R-VDR	P-VDR	TSRA	EERP
Node density 100				
2	0.01762	0.01662	0.01812	0.02012
4	0.01770	0.01710	0.01835	0.02035
6	0.01775	0.01724	0.0187	0.0207
8	0.01790	0.01745	0.0193	0.02083
10	0.01801	0.01769	0.0199	0.0213
12	0.01821	0.01778	0.0210	0.0224
14	0.01845	0.01799	0.0215	0.0245
Node density 200				
2	0.01752	0.0162	0.01810	0.0182
4	0.01754	0.0162	0.0180	0.0185
6	0.01770	0.0166	0.0194	0.0190
8	0.01785	0.0169	0.0197	0.0195
10	0.01791	0.0172	0.0204	0.0200
12	0.0182	0.0176	0.0210	0.0210
14	0.0186	0.0180	0.0215	0.0220

Table 3 End to end transmission delay (in seconds)

Data type	Node density	R-VDR	P-VDR	TSRA	EERP
TEXT	100	0.59	0.30	0.625	0.305
	200	0.52	0.291	0.580	0.298
AUDIO	100	0.697	0.357	0.621	0.305
	200	0.671	0.345	0.590	0.295
VIDEO	100	0.701	0.369	0.623	0.308
	200	0.679	0.340	0.598	0.292

from highly connected nodes to less connected nodes while ensuring shortest path. This in turn leads to more uniform distribution of load and energy consumption.

Table 3 shows the end-to-end delay experienced by text, audio and video packets respectively and Fig. 5a and b show overall value of end-to-end delay experienced by data packets from source to destination. The end-to-end delay includes both the average of route discovery time and data transmission delay from source to destination node. It can be observed that for EERP and TSRA the end-to-end delay for text, video and audio has same value, for P-VDR and R-VDR. The end-to-end delay for video and audio is comparatively higher than text messages due to adaption of paths with less locally connected edges. However, from the results (Fig. 5a, b) it has also been observed that overall average end-to-end delay for P-VDR is lower than R-VDR, EERP and TSRA. The lower delay value is due to low setup delay, lesser contention and congestion losses.

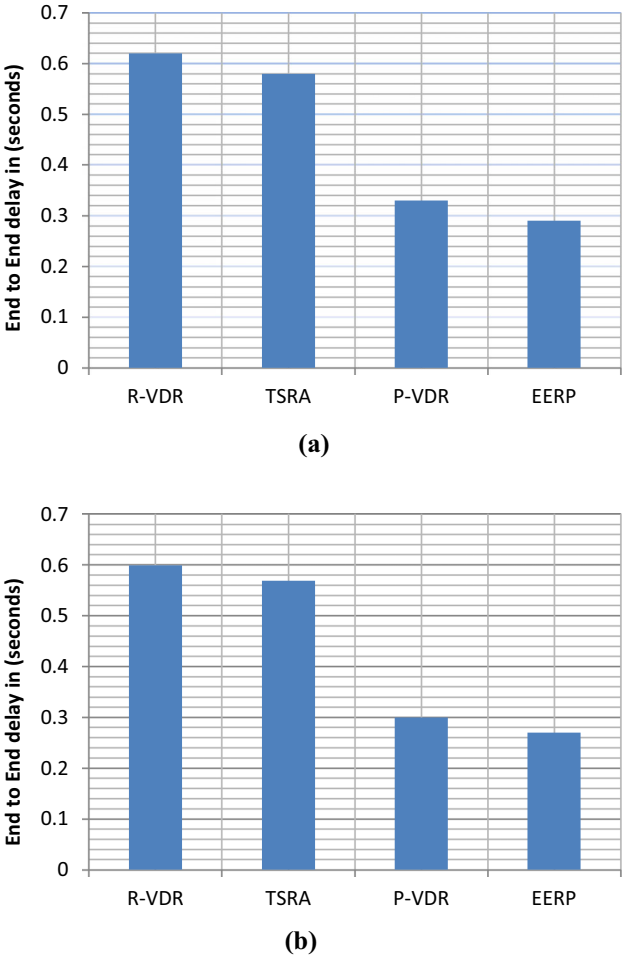


Fig. 5 Comparison of End to End delay with varying number of sensor nodes **a** N = 100, **b** N = 200

The simulation results indicate that both R-VDR and P-VDR perform better than other state-of-the-art energy efficient reactive and proactive routing protocols in terms of network lifetime, energy efficiency, connectivity percentage, packet delivery ratio and load distribution. The same can be attributed to the balanced load distribution, data type based path selection and preservation of highly connected nodes at the start of the network lifetime.

5 Conclusions and future work

Among other parameters, the efficiency of any routing protocol also varies based on the data type i.e. whether the data is text, audio or video. Routing protocols can be basically divided into two subdivisions i.e. route selection strategy and route maintenance strategy. In this paper, we have proposed a network connectivity based route selection method for different types of data viz. text, audio and video. In the proposed method, network edges are classified

on the basis of their respective network connectivity share and then the route is identified in a way so as to maximize network lifetime. Simulation results demonstrate the effectiveness of the new approach in terms of network lifetime, network connectivity, packet delivery ration, end to end delay and average energy consumption by sensor nodes. Moreover, the simulation results indicate that the proposed route selection method is not restricted to a particular route exploration strategy and can be used for developing reactive, proactive, multipath and cluster based routing. In future, the proposed route selection method can be extended for chain based routing protocols. In chain based routing protocols it is crucial to select nodes which can efficiently collect data from their neighbouring nodes as well as form an QoS aware and power efficient data routing path to deliver data to their final destination. Likewise, in the proposed method, route selection is not merely dependent upon the shortest path and has considered node density for route selection, thus it can add more efficiency to the chain based routing protocols.

References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. **52**(12), 2292–2330 (2008)
2. Lewis, F.L.: Wireless sensor networks. Smart Environ. **11–46**, 11–46 (2004)
3. Akkaya, K., Younis, M.: A survey on routing protocols for wireless sensor networks. Ad Hoc Netw. **3**(3), 325–349 (2005)
4. Jain, A.: Betweenness centrality based connectivity aware routing algorithm for prolonging network lifetime in wireless sensor networks. Wirel. Netw. **22**, 1–20 (2015)
5. Xue, Y., Cui, Y., Nahrstedt, K.: A utility-based distributed maximum lifetime routing algorithm for wireless networks. In: Proceedings of the Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QSHINE'05), pp. 10. IEEE (2005)
6. Li, Q., Aslam, J., Rus, D.: Online power-aware routing in wireless ad-hoc networks. In: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, pp. 97–107. ACM (2001)
7. Madan, R., Lall, S.: Distributed algorithms for maximum lifetime routing in wireless sensor networks. IEEE Trans. Wirel. Commun. **5**(8), 2185–2193 (2006)
8. Misra, A., Banerjee, S.: MRPC: Maximizing network lifetime for reliable routing in wireless environments. In: Proceedings of the Wireless Communications and Networking Conference, WCNC2002, vol. 2, pp. 800–806. IEEE (2002)
9. Shah-Mansouri, V., Wong, V.W.: Distributed maximum lifetime routing in wireless sensor networks based on regularization. In: Proceedings of the IEEE GLOBECOM 2007-IEEE Global Telecommunications Conference, pp. 598–603. IEEE (2007)
10. Aslam, J., Li, Q., Rus, D.: Three power-aware routing algorithms for sensor networks. Wirel. Commun. Mob. Comput. **3**(2), 187–208 (2003)
11. Akkaya, K., Younis, M.: Energy-aware delay-constrained routing in wireless sensor networks. Int. J. Commun. Syst. **17**(6), 663–687 (2004)
12. Minhas, M.R., Gopalakrishnan, S., Leung, V.C.: An online multipath routing algorithm for maximizing lifetime in wireless sensor networks. In: Proceedings of the Sixth International Conference on Information Technology: New Generations, ITNG'09, pp. 581–586. IEEE (2009)
13. Li, W., Chen, M., Li, M.M.: An enhanced aodv route protocol applying in the wireless sensor networks. InL Proceedings of the Fuzzy Information and Engineering, vol. 2, pp. 1591–1600. Springer, Berlin (2009)
14. Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: Proceedings of the WMCSA'99 Second IEEE Workshop on Mobile Computing Systems and Applications, pp. 90–100. IEEE (1999)
15. Heo, J., Hong, J., Cho, Y.: EARQ: Energy aware routing for real-time and reliable communication in wireless industrial sensor networks. IEEE Trans. Ind. Inform. **5**(1), 3–11 (2009)
16. Singh, S., Woo, M., Raghavendra, C.S.: Power-aware routing in mobile ad hoc networks. In: Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pp. 181–190. ACM (1998)
17. Park, J., Sahni, S.: An online heuristic for maximum lifetime routing in wireless sensor networks. IEEE Trans. Comput. **55**(8), 1048–1056 (2006)
18. Orojloo, H., Haghighat, A.T.: A Tabu search based routing algorithm for wireless sensor networks. Wirel. Netw. **22**, 1–14 (2016)
19. Chang, J.H., Tassiulas, L.: Maximum lifetime routing in wireless sensor networks. IEEE/ACM Trans. Netw. (TON) **12**(4), 609–619 (2004)
20. Akkaya, K., Younis, M.: An energy-aware QoS routing protocol for wireless sensor networks. In: Proceedings. 23rd International Conference on Distributed Computing Systems Workshops, pp. 710–715. IEEE (2003)
21. Ben-Othman, J., Yahya, B.: Energy efficient and QoS based routing protocol for wireless sensor networks. J. Parallel Distrib. Comput. **70**(8), 849–857 (2010)
22. Mohajerani, A., Gharavian, D.: An ant colony optimization based routing algorithm for extending network lifetime in wireless sensor networks. Wirel. Netw. **22**(8), 1–11 (2015)
23. Ganesan, D., Govindan, R., Shenker, S., Estrin, D.: Highly-resilient, energy-efficient multipath routing in wireless sensor networks. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(4), 11–25 (2001)
24. Lou, W.: An efficient N-to-1 multipath routing protocol in wireless sensor networks. In: Proceedings of the IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, pp. 8. IEEE (2005)
25. Lu, M.Y., Wong, W.S.V.: An energy-efficient multipath routing protocol for wireless sensor networks. Int. J. Commun. Syst. **20**(7), 747–766 (2007)
26. Boulfekhar, S., Benmohammed, M.: A novel energy efficient and lifetime maximization routing protocol in wireless sensor networks. Wirel. Personal Commun. **72**(2), 1333–1349 (2013)
27. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, p. 10. IEEE (2000)
28. Watfa, M.K., Mirza, O., Kawtharani, J.: BARC: a Battery Aware Reliable Clustering algorithm for sensor networks. J. Netw. Comput. Appl. **32**(6), 1183–1193 (2009)
29. Khan, Z.A., Sampalli, S.: AZR-LEACH: an energy efficient routing protocol for wireless sensor networks. Int. J. Commun. Netw. Syst. Sci. **5**(11), 785 (2012)
30. Tyagi, S., Gupta, S.K., Tanwar, S., Kumar, N.: EHE-LEACH: Enhanced heterogeneous LEACH protocol for lifetime enhancement of wireless SNs. In: Proceedings of the 2013 International

Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1485–1490. IEEE (2013)
31. Manjeshwar, A., Agrawal, D.P.: TEEN: a routing protocol for enhanced efficiency in wireless sensor networks, vol. 1, p. 189. IPDPS (2013)
32. Manjeshwar, A., Agrawal, D.P.: APTEEN: a hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks, vol. 2, p. 48. IPDPS (2002)
33. Lindsey, S., Raghavendra, C.S.: PEGASIS: Power-efficient gathering in sensor information systems. In: Proceedings of the Aerospace Conference, vol. 3, pp. 3–1125. IEEE (2002)
34. Ye, F., Luo, H., Cheng, J., Lu, S., Zhang, L.: A two-tier data dissemination model for large-scale wireless sensor networks. In: Proceedings of the 8th Annual International Conference on Mobile Computing and Networking, pp. 148–159. ACM (2002)
35. Pradittasnee, L., Camtepe, S., Tian, Y.-C.: Efficient route update and maintenance for reliable routing in large-scale sensor networks. IEEE Trans. Ind. Inform. **13**(1), 144–156 (2017)
36. Smaragdakis, G., Matta, I., Bestavros, A.,, et al.: Sep: a stable election protocol for clustered heterogeneous wireless sensor networks. In: Proceedings of the Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004), pp. 1–11. (2004)
37. Malluh, A.A., Elleithy, K.M., Qawaqneh, Z., Mstafa, R.J., Alanazi, A.: Em-sep: an efficient modified stable election protocol. In: Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1), pp. 1–7. (2014)
38. Mohemed, R.E., Saleh, A.I., Abdelrazzak, M., Samra, A.S.: Energy-efficient routing protocols for solving energy hole problem in wireless sensor networks. Comput. Netw. **114**, 51 (2016). https://doi.org/10.1016/j.comnet.2016.12.011
39. Jain, A., Reddy, B.R.: Eigenvector centrality-based cluster size control in randomly deployed wireless sensor networks. Expert Syst. Appl. **42**(5), 2657–2669 (2015)
40. Heinzelman, W.R., Kulik, J., Balakrishnan, H.: Adaptive protocols for information dissemination in wireless sensor networks. In: Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pp. 174–185. ACM (1999)
41. Jain, A., Reddy, B.R.: Ant colony optimization based orthogonal directional proactive–reactive routing protocol for wireless sensor networks. Wirel. Personal Commun. **85**(1), 179–205 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Aarti Jain an Assistant Professor in Ambedkar Institute of Advanced Communication Technology and Research, Under Govt. Of NCT Delhi, affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. She holds a Ph.D. in Communications technologies from Guru Gobind Singh Indraprastha University, Delhi, India and She received her Master's degree in Electronics and communications from Delhi College of Engineering, University of Delhi, India in 2009. Her research interests are in Energy and quality management for wireless sensor networks, bio inspired computing and its application, localization Technologies. She has more than 20 published papers in refereed National/International Journals & Conferences (viz. IEEE, ACM, Springer, Inderscience, and Elsevier) and is reviewers of various national and international reputed journals.



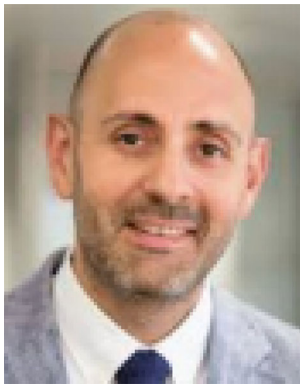
Manju Khari an Assistant Professor in Ambedkar Institute of Advanced Communication Technology and Research, Under Govt. of NCT Delhi affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. She is also the Professor- In-charge of the IT Services of the Institute and has experience of more than twelve years in Network Planning & Management. She holds a Ph.D. in Computer Science & Engineering from National Institute of Technology Patna and She received her master's degree in Information Security from Ambedkar Institute of Advanced Communication Technology and Research, formally this institute is known as Ambedkar Institute of Technology affiliated with Guru Gobind Singh Indraprastha University, Delhi, India. Her research interests are software testing, software quality, software metrics, information security, optimization and nature-inspired algorithm. She has 70 published papers in refereed National/International Journals & Conferences (viz. IEEE, ACM, Springer, Inderscience, and Elsevier), 06 book chapters in a springer. She is also co-author of two books published by NCERT of Secondary and senior Secondary School.



Elena Verdú received her master's and Ph.D. degrees in telecommunications engineering from the University of Valladolid, Spain, in 1999 and 2010, respectively. She is currently an Associate Professor at Universidad Internacional de La Rioja (UNIR) and member of the Research Group "Data Driven Science" of UNIR. For more than 15 years, she has worked on research projects at both national and European levels. Her research has focused on e-learning technologies, intelligent tutoring systems, competitive learning systems, accessibility, data mining and expert systems.



Shigeru Omatsu (M'76) was born on December 16, 1946. He received the B.E. degree in electrical engineering from Ehime University, Matsuyama, Japan, in 1969, and the M.E. and Ph.D. degrees in electronics engineering from Osaka Prefecture University, Sakai, Japan, in 1971 and 1974, respectively., He had been with Tokushima University as a Research Associate since 1974, a Lecturer since 1975, an Associate Professor since 1980, and a professor since 1988. Then, he joined Osaka Prefecture University in 1995. He is currently a Professor in the Department of Electronics, Information and Communication Engineering, Osaka Institute of Technology, Osaka, Japan. He is the author of Neuro-Control and its Applications. Berlin, Germany: Springer, and Distributed Parameter Systems: Theory and Applications. New York: Oxford Univ. Press, 1989.,Dr. Omatu has served as an Editor for the Institute of System, Control and Information Engineers in 1998 and IEE of Japan in 1999. He was an Associate Editor of the IEEE Transaction on Neural Networks (1994–1998), IMA Journal of Mathematical Control and Information (Oxford University Press), Engineering Applications of Artificial Intelligence (Pergamon), etc. He also received the Best Paper Awards from IEE of Japan and Society of Instruments and Control Engineers, in 1991 and 1995, respectively. (Based on document published on 12 September 2011).



Rubén González Crespo has a PhD in Computer Science Engineering. Currently he is Vice Chancellor of Academic Affairs and Faculty from UNIR and Global Director of Engineering Schools from PROEDUCA Group. He is advisory board member for the Ministry of Education at Colombia and evaluator from the National Agency for Quality Evaluation and Accreditation of Spain (ANECA). He is member from different committees at ISO Organization. Finally he has published more than 200 paper in indexed journals and congresses.

Theory in Biosciences (2020) 139:225–234
https://doi.org/10.1007/s12064-020-00314-6

ORIGINAL ARTICLE

A study on the stability behavior of an epidemic model with ratio-dependent incidence and saturated treatment

Abhishek Kumar[1] · Manoj Kumar[2] · Nilam[1]

Received: 10 October 2019 / Accepted: 7 March 2020 / Published online: 30 March 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
In the present article, the dynamics of a novel combination of ratio-dependent incidence rate and saturated treatment rate in susceptible-infected-recovered disease compartmental model has been presented. The ratio-dependent incidence rate has been incorporated into the model to monitor the situation when ratio of the number of infectives to that of the susceptibles is getting higher. The saturated treatment rate of the infected population has been considered as Holling type II functional, which explains the limitation in treatment availability. From the mathematical analysis of the model, two types of equilibria of the model have been obtained, which are named as disease-free equilibrium (DFE) and endemic equilibrium (EE). The local stability behavior of equilibria has been investigated by the basic reproduction number $\left(R_0\right)$, center manifold theory and Routh–Hurwitz criterion. It has been investigated that the DFE is locally asymptotically stable when $R_0 < 1$, and when $R_0 = 1$, the DFE exhibits either a forward bifurcation or a backward bifurcation under some conditions. The local stability behavior of the EE has also been analyzed, and some conditions are obtained for the same. Finally, some numerical computations have been performed in support of our theoretical results.

Keywords Epidemic model · Ratio-dependent incidence rate · Holling type II treatment rate · Bifurcations · Stability

Mathematical Subject Classification 34D20 · 92B05

Introduction

The mathematical disease transmission models are studying from the pioneering work of Kermack and Mckendrick (1927). They proposed the disease transmission model with a bilinear incidence rate (βSI). The incidence rate is defined as "the number of new cases of a disease within a period, as a proportion of the number of people at risk for the disease." The bilinear incidence rate captures the dynamics of the disease of the short amount of population; it is not reasonable for a large population because if the number of susceptibles increases, then the number of infected individuals per unit time also increases. Therefore, to understand the disease dynamics of new infection for a large population, several models have been developed by many researchers with nonlinear incidence rate such as SIR, SIRS, SEIR SVEIR, SEIRS, etc. (where S, V, E, I and R represent the susceptible, vaccinated, exposed, infected and recovered populations) (Kumar et al. 2019; Kumar and Nilam 2019a, b; Dubey et al. 2013, 2015; Li and Zhang 2017; Huang et al. 2010; Goel and Nilam 2019; Zhou and Fan 2012; Dubey et al. 2016; Xu and Ma 2009; Wang 2002; Liu et al. 1986). Capasso and Serio (1978) introduced the saturated incidence rate $g(I)S$, with

$$g(I) = \frac{kI}{1 + \alpha I}$$

where k is a positive constant and α is a nonnegative constant. In this incidence rate, kI measures the infection force of disease and $\frac{1}{1+\alpha I}$ measures the inhibition effect. Note that, the infection force $g(I)$ of classical disease transmission models is typically only a function of infected individuals.

✉ Nilam
nilam.dtu17@gmail.com; rathi.nilam@gmail.com

Abhishek Kumar
abhishek.dtu14@gmail.com

Manoj Kumar
punia.manoj@gmail.com

[1] Department of Applied Mathematics, Delhi Technological University, Delhi 110042, India

[2] Institute of System Studies and Analysis, DRDO, Metcalfe House, Delhi 110054, India

But in the transmission of communicable diseases, it involves both infective individuals and susceptible individuals. Therefore, in 2010, Li et al. (Li et al. 2011) introduced the ratio-dependent incidence rate in which infectious force is ought to depend on the number of both infective individuals and susceptible individuals of the following form:

$$g(I,S) = g\left(\frac{I}{S}\right) = \frac{k\left(\frac{I}{S}\right)}{1+\alpha\left(\frac{I}{S}\right)}$$

where α is a nonnegative constant. The ratio-dependent incident rate $\frac{k\left(\frac{I}{S}\right)}{1+\alpha\left(\frac{I}{S}\right)}S$ means that the rate of spread of the disease is approximately controlled by kI in the early stages of disease transmission (e.g., $\frac{I}{S}$ is small), and it is approximately controlled by $\frac{k}{\alpha}S$ in the endemic when almost everyone is infected, in other words, when $\frac{I}{S}$ is large. Thus, the ratio-related incidence of transmission takes into account inhibition effects and behavioral changes during epidemics.

The treatment of infectious disease contributes greatly to cure infectives and interrupt transmission of infection to susceptibles. The appropriate and timely treatment strategy can significantly lessen the effect of disease on society. In classical epidemic models, the treatment rate of the infected individuals is assumed to be either constant or proportional to the number of infected people. However, we know that there are limited treatment resources available in the community. Therefore, it is of grave importance to choose a suitable treatment rate of a disease. Zhou and Fan (2012) proposed an epidemic model with the continuous, differentiable nonlinear treatment rate, known as the Holling type II treatment rate as follows:

$$T(I) = \frac{aI}{1+bI}, \quad I \geq 0, \quad a > 0, \quad b \geq 0.$$

They have shown that with varying amount of medical resources and their supply efficiency, the target model admits both backward and Hopf bifurcation. In Holling type II treatment rate, for any outbreak of the disease, its treatment capacity is first very low and then grows slowly with the improvement in hospital's condition, availability of effective drugs, etc. Further, when the number of infected individuals is very large, the treatment capacity reaches its maximum due to limited treatment facilities. This condition pertains to newly emergent diseases whose treatment is very limited.

The objective of this article is to explore the disease transmission dynamics through the SIR model when the ratio of the number of infectives to that of the susceptibles is getting larger, and the treatment facilities for the infected individuals are limited. For this purpose, we propose a SIR model with the combination of the ratio-dependent incidence rate $g(S,I)S$ and Holling type II ($T(I)$) treatment rate. The dynamical behavior of the model is analyzed by the basic reproduction number R_0, Routh–Hurwitz criterion and center manifold theory. We investigate the result at the critical value of R_0 for the elimination/persistence of the disease by proving the results for the forward and backward bifurcations under some conditions. We also investigate the result for the reappearance of the infection by the occurrence of limit cycles and Hopf bifurcation.

The rest of the article is organized as follows: A mathematical model is proposed in Sect. 2. The basic properties of the model are discussed in Sect. 3. In Sect. 4, we obtain the model equilibria and discuss their stability behaviors. The model is simulated numerically in Sect. 5. Finally, a discussion is given in Sect. 6.

Mathematical model

In this section, we present a mathematical model for disease transmission under the assumption that the total population size N is fixed (or constant). We divide the total constant population into three disjoint subpopulations (or compartments), namely susceptible $S(t)$, infected $I(t)$ and recovered $R(t)$. It is supposed that the disease can be transmitted to a susceptible individual only by close contact with an infected individual. It is also supposed that a recovered individual will not be reinfected in the future. Let A denote the constant recruitment rate of susceptibles, μ be the natural death rate of the population, α be the transmission rate of infection, β be the rate of inhibition from the behavioral changes of the susceptibles, d be the rate of disease-induced mortality, θ be the natural recovery rate of infected individuals, a be the cure rate to infection and b be the limitation rate in treatment availability of infected individuals. The transfer diagram of an epidemic through various compartments is shown in Fig. 1.

Under the above assumptions, the disease transmission model is described by the following system of nonlinear ordinary differential equations:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= A - \mu S - g(S,I)S,\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= g(S,I)S - (\mu + d + \theta)I - T(I),\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= T(I) + \theta I - \mu R.
\end{aligned} \tag{1}$$

where $S(0) > 0, I(0) \geq 0$, $R(0) \geq 0$, $g(S,I) = \frac{\alpha I}{S+\beta I}$ and $T(I) = \frac{aI}{1+bI}$.

The description of the symbols of the model is given in Table 1.

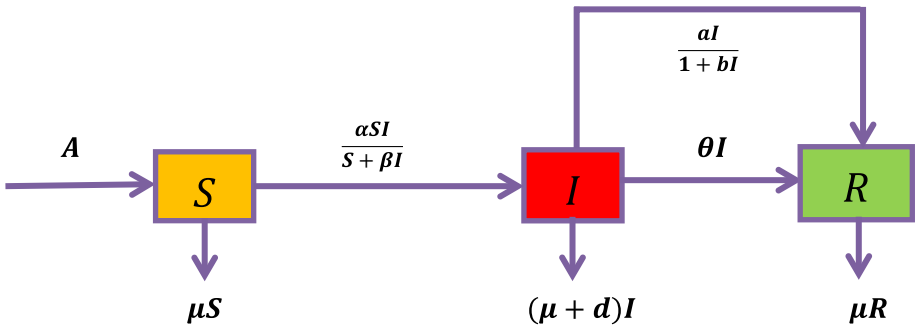


Fig. 1 Transfer diagram of infection from susceptible (S) individuals through infected (I) and recovered (R) compartments for the model

Table 1 Description of symbols

Symbol	Description
S	Susceptible population
I	Infected population
R	Recovered population
A	Constant recruitment rate
μ	Natural death rate
α	Transmission rate
β	Inhibition effects
d	Disease-induced death rate
θ	Recovery rate
a	Cure rate
b	Limitation rate in treatment availability
R_0	Basic reproduction number
$g(S,I)S$	Ratio-dependent incidence rate
$T(I)$	Holling type II treatment rate

Basic properties

For biological reasons, it is assumed that all the state variables (S, I, R) and the parameters $(A, \beta, \alpha, \mu, d, \theta, a, b)$ of the model (1) are nonnegative. Now, we show that the model solutions always enter in a compact region $D_* = \left\{(S,I,R) \in \mathbb{R}^3_+ : S+I+R \le \frac{A}{\mu}\right\}$. It can be shown as follows:

Note that

$$\frac{\mathrm{d}S}{\mathrm{d}t} + \frac{\mathrm{d}I}{\mathrm{d}t} + \frac{\mathrm{d}R}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(S+I+R)$$
$$\frac{\mathrm{d}}{\mathrm{d}t}(S+I+R) = \frac{\mathrm{d}N}{\mathrm{d}t} = 0 \quad (\therefore S+I+R=N)$$
$$\frac{\mathrm{d}S}{\mathrm{d}t} + \frac{\mathrm{d}I}{\mathrm{d}t} + \frac{\mathrm{d}R}{\mathrm{d}t} = 0$$
$$A - \mu N - \mathrm{d}I = 0$$
$$\mu N + \mathrm{d}I = A$$
$$N \le \frac{A}{\mu}.$$

Therefore,

$$\lim_{t\to\infty}(S(t)+I(t)+R(t)) \le \frac{A}{\mu}.$$

It follows that all the nonnegative solutions of the model (1) approach inside the invariant set

$$D_* = \left\{(S,I,R) \in \mathbb{R}^3_+ : S+I+R \le \frac{A}{\mu}\right\}.$$

Hence, we propose the following lemma:

Lemma 1 *The set* $D_* = \left\{(S,I,R) \in \mathbb{R}^3_+ : S+I+R \le \frac{A}{\mu}\right\}$ *is an invariant manifold of the model* (1)*, which is attracting the first octant.*

It is noticed that the right hand of the model (1) is completely continuous and locally Lipschitzian on D_*. Thus, it follows that the solution of the model (1) with nonnegative initial conditions exists and is unique. Hence, the model (1) is mathematically and epidemiologically well-posed.

Under the assumption that the population size N is constant, the model (1) may be reduced to a two-dimensional system. The third equation (equation of R) of the model (1) is traditionally omitted; it is easy to see that under the assumption of constant population size, this equation is decoupled from the first two equations of the model (1). The condition $N = S(t) + I(t) + R(t) =$ Constant may be used to find $R(t)$. Therefore, following reduced system is considered for further study:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= A - \mu S - \frac{\alpha SI}{S+\beta I},\\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\alpha SI}{S+\beta I} - (\mu + d + \theta)I - \frac{aI}{1+bI}.\end{aligned} \tag{2}$$

Equilibria and their stability analysis

In this section, we analyze system (2) by finding its equilibria and investigating stability analysis. We find that the system (2) has two equilibria which are obtained by setting right-hand sides terms of the system (2) to zero, given as:

i. Disease-free equilibrium $Q\left(\frac{A}{\mu},0\right)$.
ii. Endemic equilibrium $Q^*(S^*,I^*)$.

Next, to explore the stability behavior of equilibria, first, we compute the basic reproduction number by the next-generation matrix method (Driessche and Watmough 2002) as below:

Suppose that

$$\dot{G} = X(y) - Y(y),$$

where $y = (I,S)^T$, $X(y)$ is the matrix of new infection term, and $Y(y)$ is the matrix of transfer term into the compartments and out of the compartments. Jacobian matrices of $X(y)$ and $Y(y)$ at $Q\left(\frac{A}{\mu},0\right)$ are given by

$$X = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} a+d+\theta+\mu & 0 \\ \alpha & \mu \end{pmatrix}.$$

Now,

$$XY^{-1} = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{a+d+\theta+\mu} & 0 \\ -\frac{\alpha}{\mu(a+d+\theta+\mu)} & \frac{1}{\mu} \end{pmatrix} = \begin{pmatrix} \frac{\alpha}{a+d+\theta+\mu} & 0 \\ 0 & 0 \end{pmatrix}$$

The basic reproduction number is the spectral radius of the next-generation matrix XY^{-1} (Driessche and Watmough 2002). Therefore, the basic reproduction number (R_0) for our model is

$$R_0 = \rho\left(XY^{-1}\right) = \frac{\alpha}{\mu+d+\theta+a}.$$

Thus, by the basic reproduction number R_0 and Theorem 2 of Driessche and Watmough (2002), we state the following theorem:

Theorem 1 *The disease-free equilibrium $Q\left(\frac{A}{\mu},0\right)$ is locally asymptotically stable if basic reproduction number R_0 is less than one and unstable if R_0 is greater than one.*

Analysis at $R_0 = 1$

In this subsection, we check the stability behavior of the disease-free equilibrium $Q\left(\frac{A}{\mu},0\right)$ when the basic reproduction number R_0 is equal to one.

Let us redefine $S = y_1$ and $I = y_2$. Then, the system (2) can be rewritten as:

$$\begin{aligned} \frac{dy_1}{dt} &= A - \mu y_1 - \frac{\alpha y_1 y_2}{y_1+\beta y_2} \equiv g_1, \\ \frac{dy_2}{dt} &= \frac{\alpha y_1 y_2}{y_1+\beta y_2} - (\mu+d+\theta)y_2 - \frac{a y_2}{1+b y_2} \equiv g_2. \end{aligned} \tag{3}$$

The linearized matrix of the system (3) at $Q\left(\frac{A}{\mu},0\right)$ and bifurcation parameter $\alpha = \alpha^* = (\mu+d+\theta+a)$ is given by

$$M = \begin{pmatrix} -\mu & -\alpha^* \\ 0 & \alpha^* - (a+d+\theta+\mu) \end{pmatrix} = \begin{pmatrix} -\mu & -\alpha^* \\ 0 & 0 \end{pmatrix}.$$

At $R_0 = 1$, the matrix M has a simple zero eigenvalue and the second eigenvalue is negative. At this stage, the linearization technique fails to conclude the behavior of the system (3) (Dubey et al. 2013; Sastry 1999). Therefore, we applied the center manifold theory to study the behavior of the equilibrium. Then, from (Chavez and Song 2004), the bifurcation constants K_1 and K_2 are given by

$$K_1 = \sum_{k,i,j=1}^{2} u_k w_i w_j \left(\frac{\partial^2 g_k}{\partial y_i \partial y_j}\right)_Q,$$

and

$$K_2 = \sum_{k,i=1}^{2} u_k w_i \left(\frac{\partial^2 g_k}{\partial y_i \partial \beta^*}\right)_Q$$

where $w = \left(w_1, w_2\right)^{\mathrm{T}}$ and $u = \left(u_1, u_2\right)$ are the right eigenvector and the left eigenvector of the matrix M corresponding to a zero eigenvalue, respectively. Then we have

$$u_1 = 0, \quad u_2 = 1 \quad \text{and} \quad w_1 = -\frac{\alpha^*}{\mu}, \quad w_2 = 1$$

The nonzero partial derivatives associated with the functions $g_i, i = 1,2$ of the system (3) evaluated at $R_0 = 1$ and $\alpha = \alpha^*$ are

$$\left(\frac{\partial^2 g_2}{\partial x_2^2}\right)_Q = 2ab - \frac{2\alpha^*\beta\mu}{A} \quad \text{and} \quad \left(\frac{\partial^2 g_2}{\partial x_2 \partial \alpha^*}\right)_Q = 1$$

Therefore,

$$K_1 = u_2\left(w_2^2 \cdot \left(2ab - \frac{2\alpha^*\beta\mu}{A}\right)\right) = 2ab - \frac{2\alpha^*\beta\mu}{A},$$

and

$$K_2 = u_2\left(w_2 . 1\right) = 1 > 0.$$

We see that the bifurcation constant K_2 is positive. Therefore, using (Chavez and Song 2004), the local behavior of the system at $R_0 = 1$ depends on the sign of bifurcation constant K_1. On evaluating K_1 at bifurcation parameter α^*, we get

$$K_1 = 2\left(ab - \frac{\beta\mu(a+d+\theta+\mu)}{A}\right) = \frac{2}{A}(a(Ab-\beta\mu) - \beta\mu(d+\theta+\mu)).$$

It can be seen that $K_1 > 0$ when $a > \varpi$, and $K_1 < 0$ when $a < \varpi$, where $\varpi = \frac{\beta\mu(d+\theta+\mu)}{Ab-\beta\mu}$.

Hence, we propose the following theorem.

Theorem 2 *The system (2) at $R_0 = 1$ exhibits the following transcritical bifurcation:*

i. *a forward bifurcation when $0 \le a < \varpi$.*
ii. *a backward bifurcation when $a > \varpi$.*

The forward and backward transcritical bifurcations are illustrated in Fig. 2a, b with parameter values given in Table 2.

From the above set of parameter values, we calculated that $\varpi = 0.048$. Figure 2a shows the existence of forward bifurcation for the cure rate $a = 0.025$, which satisfies Theorem 2(i). Figure 2b shows the occurrence of backward bifurcation when the cure rate is $a = 0.06$, which satisfies Theorem 2(ii).

Existence of the endemic equilibrium

To find the conditions for the existence of endemic equilibrium $Q^*(S^*, I^*)$, the system (2) is rearranged to get S^* and I^* which gives:

i. $S^* = -\frac{I^*\beta(a+(1+bI^*)(d+\theta+\mu))}{a+(1+bI^*)(d-\alpha+\theta+\mu)}$.
ii. I^*, given by the equation

$$C_3 I^{*3} + C_2 I^{*2} + C_1 I^* + C_0 = 0, \tag{4}$$

where

$$\begin{aligned}
C_0 &= A(a+d+\theta+\mu-\alpha) = A(\mu+d+\theta+a)\left(1-R_0\right),\\
C_1 &= -a^2 + 2Ab(d-\alpha+\theta+\mu) + a(Ab-2d+\alpha-2\theta+(-2+\beta)\mu)\\
&\quad -(d+\theta+\mu)(d-\alpha+\theta+\mu-\beta\mu),\\
C_2 &= b(Ab(d-\alpha+\theta+\mu) + a(-2d+\alpha-2\theta+(-2+\beta)\mu) - 2(d+\theta+\mu)(d-\alpha+\theta+\mu-\beta\mu)),\\
C_3 &= -b^2(d+\theta+\mu)(d-\alpha+\theta+\mu-\beta\mu).
\end{aligned}$$

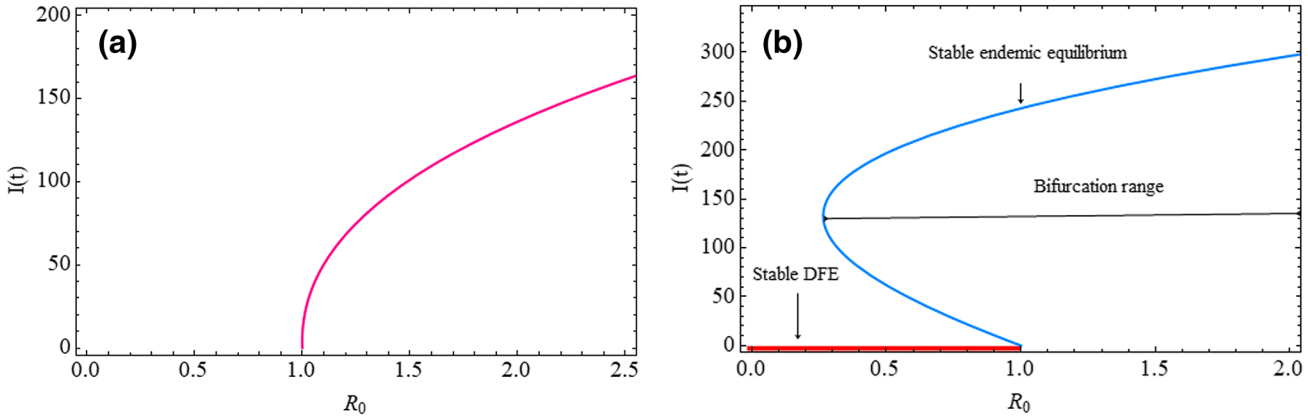


Fig. 2 **a** Forward bifurcation graph with $a = 0.025$. **b** Backward bifurcation graph with $a = 0.06$

Table 2 Parameters and their numerical values

Parameter	A	μ	α	d	θ	β	b
Numerical value	3	0.03	0.05	0.001	0.001	0.3	0.005

For $R_0 > 1$, using the Descartes' rule of signs (Kermack and McKendrick 1927), Eq. (4) has a unique positive real root I^* if anyone from the following holds:

$C_3 > 0, \quad C_2 > 0, \quad C_1 > 0, \quad C_0 < 0.$
$C_3 > 0, \quad C_2 > 0, \quad C_1 < 0, \quad C_0 < 0.$
$C_3 > 0, \quad C_2 < 0, \quad C_1 < 0, \quad C_0 < 0.$

After obtaining the value of I^*, we can get the value of S^*. It implies that the system (2) has a unique endemic equilibrium $Q^*(S^*, I^*)$ if anyone of the above conditions holds.

Next, we explore the stability behavior of the endemic equilibrium $Q^*(S^*, I^*)$.

Stability analysis of the endemic equilibrium

To explore the stability behavior of the endemic equilibrium $Q^*(S^*, I^*)$, we obtain a variational matrix corresponding to $Q^*(S^*, I^*)$ of the system (2) as below:

$$M^*_{Q^*(S^*,I^*)} = \begin{pmatrix} \frac{S^*I^*\alpha}{(S^*+I^*\beta)^2} - \frac{I^*\alpha}{S^*+I^*\beta} - \mu & -\frac{S^{*2}\alpha}{(S^*+I^*\beta)^2} \\ \frac{I^{*2}\alpha\beta}{(S^*+I^*\beta)^2} & -d - \frac{a}{(1+bI^*)^2} - \frac{S^*I^*\alpha\beta}{(S^*+I^*\beta)^2} + \frac{S^*\alpha}{S^*+I^*\beta} - \theta - \mu \end{pmatrix}$$

The characteristic equation of the matrix $M^*_{Q^*(S^*,I^*)}$ is given by the following equation

$$\tau^2 + x_*\tau + y_* = 0 \tag{5}$$

where

$x_* = x_a - x_b, y_* = y_a - y_b,$

with

$$x_a = d + \frac{a}{(1+bI^*)^2} + \frac{I^{*2}\alpha\beta}{(S^*+I^*\beta)^2} + \theta + 2\mu, \quad x_b = \frac{S^{*2}\alpha}{(S^*+I^*\beta)^2},$$
$$y_a = \frac{\left(a + (1+bI^*)^2(d+\theta+\mu)\right)\left(S^{*2}\mu + 2S^*I^*\beta\mu + I^{*2}\beta(\alpha+\beta\mu)\right)}{(1+bI^*)^2(S^*+I^*\beta)^2}, \quad y_b = \frac{S^{*2}\alpha\mu}{(S^*+I^*\beta)^2}.$$

Now, we apply the Routh–Hurwitz criterion and propose the following theorems:

Theorem 3 *Endemic equilibrium $Q^*(S^*, I^*)$ is*

i. *locally asymptotically stable when the conditions $x_a > x_b$ and $y_a > y_b$ simultaneously hold.*
ii. *a saddle point when the conditions $x_a > x_b$ and $y_a < y_b$ simultaneously hold.*
iii. *unstable when the conditions $x_a < x_b$ and $y_a > y_b$ simultaneously hold.*

Theorem 4 *The system (2) exhibits a Hopf bifurcation near the endemic equilibrium $Q^*(S^*, I^*)$ if the conditions $y_a > y_b$ and $x_a = x_b$ hold simultaneously.*

Proof The condition $x_a = x_b$ implies that $x_* = 0$ in Eq. (5) and the condition $y_a > y_b$ implies that $y_* > 0$. Therefore, Eq. (5) has purely imaginary roots. From Theorems 3(*i*) and 4(*ii*), it follows that the equilibrium $Q^*(S^*, I^*)$ changes its behavior from stable to instability as the parameter α passes through the critical value $\alpha = \alpha^*$, where

$$\alpha^* = -\frac{(S^*+I^*\beta)^2\left(a + (1+bI^*)^2(d+\theta+2\mu)\right)}{(1+bI^*)^2\left(-S^{*2}+I^{*2}\beta\right)}$$

Again, we have

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\left[tr\left(M_{Q^*}\right)\right]_{\alpha=\alpha^*} = -\frac{I^{*2}\beta - S^{*2}}{(S^*+I^*\beta)^2} = \frac{1}{\alpha^*}\left(\frac{a + (1+bI^*)^2(d+\theta+2\mu)}{(1+bI^*)^2}\right) \neq 0.$$

Hence, the system (2) exhibits a Hopf bifurcation near the equilibrium point $Q^*(S^*, I^*)$ when $\alpha = \alpha^*$.

Theorem 5 *If $a \leq 2\mu$, then the system (2) does not have any periodic solution in the interior of the positive quadrant of the S–I plane.*

Proof We define a real-valued function in the interior of the *S–I* plane as follows:

$$U(S, I) = \frac{S + \beta I}{SI}$$

Let us consider

Table 3 Parameters and their numerical values

Parameter	A	μ	α	d	θ	β	b	a
Numerical value	3	0.03	0.05	0.005	0.001	0.3	0.005	0.002

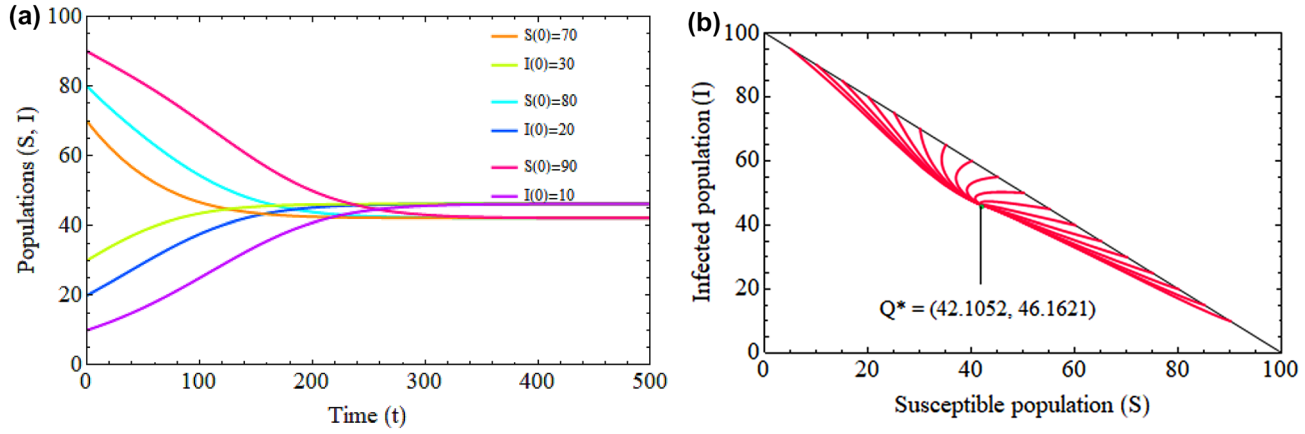


Fig. 3 **a** Susceptible and infected populations at different initial values of subpopulations. **b** Local asymptotical stability behavior of an endemic equilibrium $Q^*(42.1052, 46.1621)$ when $R_0 = 1.3157 (R_0 > 1)$

$$g_1(S,I) = A - \mu S - \frac{\alpha SI}{S+\beta I},$$
$$g_2(S,I) = \frac{\alpha SI}{S+\beta I} - (\mu + d + \theta)I - \frac{aI}{1+bI}.$$

Then, we have

$$\operatorname{div}\left(Ug_1, Ug_2\right) = \frac{\partial}{\partial S}\left(Ug_1\right) + \frac{\partial}{\partial I}\left(Ug_2\right)$$
$$= -\frac{AY(1+bI)^2\beta + SI\beta\left(a + (1+bI)^2(d+\theta)\right) + S\left(S + b^2SI^2 + I(1+bI)^2\beta\right)\mu}{I(S+bSI)^2} + \frac{b(a-2\mu)}{(1+bI)^2}.$$

It can be seen that the above expression can never be equal to zero and also it will not change the sign in the positive quadrant of the S–I plane if the condition $a \leq 2\mu$ holds. Then, by Dulac's criterion (Sastry 1999), it can be stated that the system (2) does not have any periodic solution in the interior of the positive quadrant of the S–I plane. Biologically, this theorem says that if the given condition holds, then the disease will not reappear in the society.

Numerical simulations

The following set of numerical experimental values of the parameters (Table 3) is taken for the simulation of the system (2):

Initial values are

$$S(0) = 70, I(0) = 30$$

Figure 3a shows the susceptible and infected populations at different values of initial subpopulations. From Fig. 3a, it can be seen that for any initial subpopulations, both the subpopulations (susceptible and infected) reach the same endemic equilibrium $Q^*(42.1052, 46.1621)$. The plot of the subpopulations with respect to infected population has been drawn in Fig. 3b when the basic reproduction number R_0 is greater than one. This figure shows that when $R_0 > 1$, for any values of initial subpopulations, the susceptible and infected subpopulations achieve the same endemic equilibrium $Q^*(42.1052, 46.1621)$.

Figure 4 portrays the behavior of susceptible and infected subpopulations when the basic reproduction number R_0 is less than one. This figure shows that when $R_0 < 1$, then for any value of initial subpopulations, the susceptible population reaches its maximum value, and the infected population becomes zero, which confirms Theorem 1.

Figures 5 and 6 show the infected population at different values of the transmission rate (α) and measures of inhibitions (β), respectively. Figure 5 shows the increment in the infected population when the value of α increases, and Fig. 6

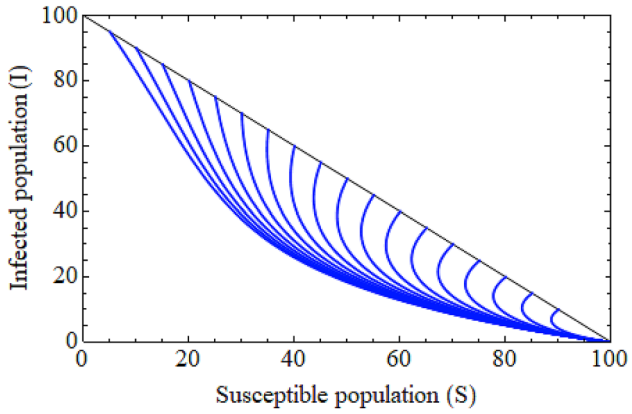


Fig. 4 Graph in support of Theorem 1 when $R_0 = 0.7042 (R_0 < 1)$ with $a = 0.035$

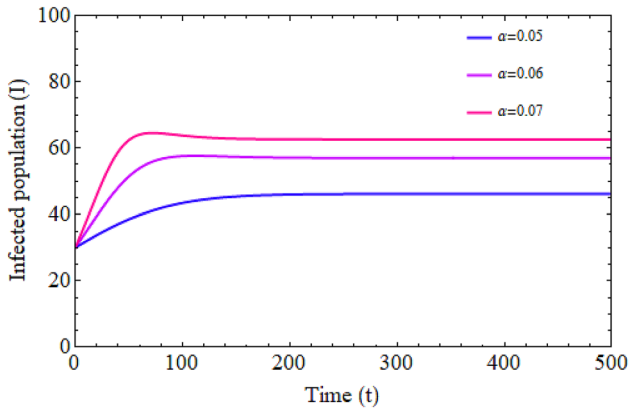


Fig. 5 Variation in the infected population at different values of the transmission rate (α)

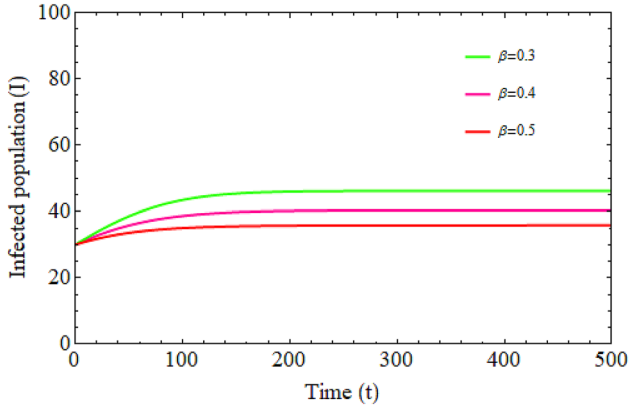


Fig. 6 Variation in the infected population at different values of inhibition effect (β)

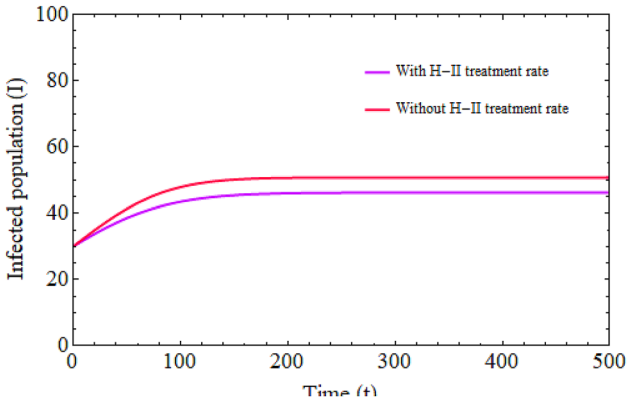


Fig. 7 Variation in the infected population with and without the Holling type II treatment rate

shows the decrement in the infected population when the value of β increases. Thus, the number of infected individuals can be reduced with the decreased transmission rate and increased inhibitions rate, which is biologically true.

Figure 7 shows the difference in the infected population with and without the Holling type II treatment rate. This figure shows that when treatment is given according to Holling type II functional to the infected population then the number of newly infected individuals becomes less in comparison with the situation that when no treatment is given to infectives. Thus, Holling type II treatment leads to decline in the prevalence of infection.

To show the numerical result in support of Theorem 5, we take the following set of parameter values (Table 4).

Figure 8 shows the stable limit cycle in the S–I plane which confirms the existence of Hopf bifurcation near the endemic equilibrium $Q^*(9.55, 2.45)$ with bifurcation parameter $\alpha^* = 1.19$.

Discussion

In the present article, the epidemiological mathematical model is meant to represent the dynamics of infectious disease when the ratio of infected population to that of susceptible population is large which takes into account: inhibition effects and behavioral changes during epidemics, and limitation in treatment facilities. For this purpose, we have considered the ratio-dependent incidence rate and Holling type II treatment rate. The stability behavior of the disease-free and the endemic equilibria is investigated, and it is obtained that the disease-free equilibrium is locally asymptotically stable when the basic reproduction number R_0 is less than unity and unstable when R_0 is greater than unity as shown analytically

Table 4 Parameters and their numerical value

Parameter	A	μ	α	d	θ	β	b	a
Numerical value	3	0.03	1.19	0.005	0.001	0.3	0.05	1.2

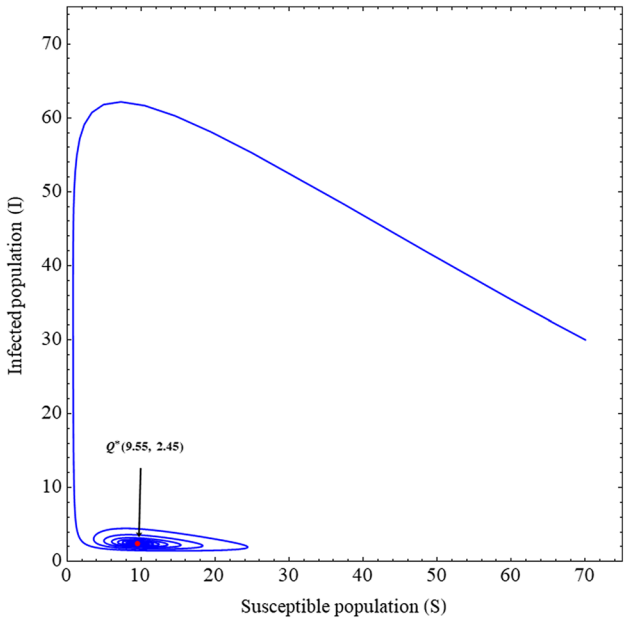


Fig. 8 Stable limit cycle in the S–I plane

and numerically in Theorem 1 and Fig. 4, respectively. We also investigated that the disease-free equilibrium at $R_0 = 1$ exhibits the forward or backward transcritical bifurcation with bifurcation parameter $\alpha = \alpha^*$ under some conditions as shown analytically in Theorem 2 and numerically in Fig. 2a, b. The occurrence of the forward and backward bifurcations has a great implication on the epidemic model. If the forward bifurcation occurs at $R_0 = 1$, then it is not possible for a disease to invade in the population when $R_0 < 1$. If the backward bifurcation occurs at $R_0 = 1$, then the classical condition to reduce the basic reproduction number below one will not be sufficient to eradicate the disease from society. We showed that the occurrence of the forward or backward bifurcation depends on the treatment given to the infected individuals. If the cure rate is suitably lower than the certain quantity ϖ (obtained in Theorem 2), then the forward bifurcation occurs, whereas if the cure rate is sufficiently high then there will be the presence of backward bifurcation, which demonstrates that the disease-free equilibrium coexists with the endemic equilibrium when basic reproduction number is less than one. Further, we showed that the endemic equilibrium is either locally asymptotically stable or a saddle point or unstable under some conditions as shown in Theorem 3. We also showed that the system (2) exhibits a Hopf bifurcation near the endemic equilibrium with bifurcation parameter $\alpha = \alpha^*$ both analytically (Theorem 4) and numerically (Fig. 8). Moreover, the nonexistence of any periodic solution in the positive quadrant of S–I plane has been shown under the condition stated in Theorem 5, which explains that the infection will not reappear in the society in future if the mentioned condition is satisfied. From numerical simulations, we observed that more effective control in the new infection can be achieved if the treatment of infectives is managed according to the Holling type II treatment rate.

Acknowledgements The authors are thankful to Delhi Technological University, Delhi, India, for monetary support for this research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Alexander ME, Bowman C, Moghadas SM, Summers R, Gumel AB, Sahai BM (2004) A vaccination model for transmission dynamics of influenza. SIAM J Appl Dyn Syst 3(4):503–524

Capasso V, Serio G (1978) A generalization of the Kermack–McKendrick deterministic epidemic model. Math Biosci 42(1–2):43–61

Chavez CC, Song B (2004) Dynamical models of tuberculosis and their applications. Math Biosci Eng 1:361–404

Driessche PVD, Watmough J (2002) Reproduction numbers and subthreshold endemic equilibria for compartment models of disease transmission. Math Biosci 180:29–48

Dubey B, Patra A, Srivastava PK, Dubey US (2013) Modelling and analysis of a SEIR model with different types of nonlinear treatment rates. J Biol Syst 21(3):1350023

Dubey B, Dubey P, Dubey US (2015) Dynamics of an SIR model with nonlinear incidence and treatment rate. Appl Appl Math 10(2):718–737

Dubey P, Dubey B, Dubey US (2016) An SIR model with nonlinear incidence rate and Holling type III treatment rate. Appl Anal Biol Phys Sci 186:63–81

Goel K, Nilam (2019) A mathematical and numerical study of a SIR epidemic model with time delay, nonlinear incidence and treatment rates. Theory Biosci. https://doi.org/10.1007/s12064-019-00275-5

Huang G, Takeuchi Y, Ma W, Wei D (2010) Global Stability for delay SIR and SEIR epidemic models with nonlinear incidence rate. Bull Math Biol 72:1192–1207

Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of Epidemics. Proc R Soc Lond A 115(772):700–721

Kumar A, Nilam (2019a) Dynamic behavior of a SIR epidemic model along with Time delay. Crowley–Martin type incidence rate and Holling type II treatment rate. Int J Nonlinear Sci Numer Simul. https://doi.org/10.1515/ijnsns-2018-0208

Kumar A, Nilam (2019b) Stability of a delayed SIR epidemic model by introducing two explicit treatment classes along with nonlinear incidence rate and Holling type treatment. Comput Appl Math 38:130

Kumar A, Goel K, Nilam (2019) A deterministic time-delayed sir epidemic model: mathematical modeling and analysis. Theory Biosci. https://doi.org/10.1007/s12064-019-00300-7

Li GH, Zhang YX (2017) Dynamic behavior of a modified SIR model in epidemic diseases using nonlinear incidence and recovery rates. PLoS ONE 12(4):e0175789

Li B, Yuan S, Zhang W (2011) Analysis on an epidemic model with a ratio-dependent nonlinear incidence rate. Int J Biomath 4(2):227–239

Liu WM, Levin SA, Iwasa Y (1986) Influence of nonlinear incidence rates upon the behavior of SIRS epidemiological models. J Math Biol 23(2):187–204

Sastry S (1999) Analysis, stability and control. Springer, New York

Wang WD (2002) Global behavior of an SEIRS epidemic model with time delays. Appl Math Lett 15:423–428

Wang X (2004) A simple proof of Descartes's rule of signs. Am Math Mon. https://doi.org/10.2307/4145072

Xu R, Ma Z (2009) Stability of a delayed SIRS epidemic model with a nonlinear incidence rate. Chaos Solut Fractals 41:2319–2325

Zhou L, Fan M (2012) Dynamics of a SIR epidemic model with limited medical resources revisited. Nonlinear Anal RWA 13:312–324

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), 28-29 Feb. 2020, Langkawi, Malaysia

An Efficient Approach to EEG-Based Emotion Recognition using LSTM Network

Anubhav
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
anubhav2901@gmail.com

Debarshi Nath
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
debarshinath94@gmail.com

Mrigank Singh
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
mriganksingh1@gmail.com

Divyashikha Sethia
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
divyashikha@dtu.ac.in

Diksha Kalra
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kalradiksha11@gmail.com

S. Indu
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
s.indu@dce.ac.in

***Abstract*—This work aims to investigate the performance of the Long Short-Term Memory (LSTM) Model for EEG-Based Emotion Recognition. For the experimentation, we use the publicly available DEAP dataset, which consists of preprocessed EEG and physiological signals. Our work limits itself to the study of only the EEG signals to have a scope for developing an efficient headgear model for real-time monitoring of emotions. In this study, we extract the band power, a frequency-domain feature, from the EEG signals and compare the classification accuracies for Valence and Arousal domain for different classifiers. The proposed Long Short-Term Memory (LSTM) model achieves the best classification accuracy of 94.69% and 93.13% for Valence and Arousal scales, respectively, illustrating a significant average increment of 16% in valence and 18% in arousal in comparison to other classifiers.**

***Keywords*—EEG Data, Emotion, Emotion Recognition, DEAP dataset, Band power, LSTM Network**

I. Introduction

Emotion represents the state of mind of a person whether a person is happy or sad, angry or calm, stressed, or relieved. Emotions are the response to a particular stimulus. Studies suggest that emotion is a subjective experience: it varies from person to person, and because of this, it is one of the most challenging and exciting research fields in psychology [1]. Recognition of emotion plays a vital role in daily life as it can help in enhancing one's psychological health which is equally important as maintaining physical fitness. Nowadays, a lot more people suffer from anxiety, stress, hypertension, and other mental health-related issues. So, Emotion Recognition here plays a crucial role in improving the lives of people. For instance, when a game becomes too dull or too exciting, the level of the game can be modified depending on the exhibited emotional level of the person. Also, a computer can change the music or window background according to one's mood. There are various other applications in the field of mental health where the knowledge of human emotion helps the psychologist to treat stress, tension, and anxiety issues.

Emotion is a phenomenon that is difficult to grasp, and for its better understanding, there are various models proposed by researchers like Valence and Arousal Model by Russell [2]. This model represents emotions on a 2-D circular space where arousal represents the vertical axis, and valence represents the horizontal axis. The Circular space represents the neutral valence and medium value of arousal. Bradley et al. [3] proposed another model named Approach and Withdrawal Model or the vector model. It is also a 2-D model where the value of valence determines the direction of emotion where a positive value of valence shifts the emotion in the top vector. Likewise, the negative value of valence would shift the emotion in the down vector. Watson and Tellegen [4] developed a Positive and Negative Model. In this model, the vertical axis represents low to high positive affect, and the horizontal axis represents low to high negative affect.

Earlier researches on emotions were done using facial expressions, speech processing, and various other methods. However, since it is possible to fake this behaviour and techniques, the focus has now shifted on emotion recognition using other physiological signals such as Electrocardiography (ECG), Electromyography (EMG), Galvanic Skin Response (GSR), Respiration Rate (RR) and Electroencephalogram (EEG) signals [5], [6]. Emotion Recognition through EEG has vast applications in the field of Human-Computer Interaction (HCI), where the computer can adjust its behaviour according to user emotion. For the measurement of brain signals, Electroencephalogram (EEG) device is used, which measures the electrical activity of the brain. EEG device contains a large number of electrodes that can be placed on the Human

978-1-7281-5310-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:03:17 UTC from IEEE Xplore. Restrictions apply.

Skull, according to the 10-20 International System. Since the understanding of the entire EEG signal at once is a complex task, the EEG signal is divided into various frequency bands. The different frequency bands are the Alpha (0-4 Hz), Beta (4-8 Hz), Delta (8-13 Hz), Theta (13-30 Hz), and Gamma (>30 Hz). Each band is associated with different activities taking place in the body. For instance, the Delta wave is related to deep sleep as well as the deepest level of relaxation. Similarly, the Theta wave is associated with REM sleep, deep and raw emotions, and cognitive processing. Likewise, in a drowsy state, the Alpha wave comes into the picture. It is associated with relaxation and calmness. In a conscious state, the Beta wave is present during the thought process. Gamma waves are current when a person tries to perceive two different senses at the same time as sound and sight. EEG signals have broad applications ranging from Emotion Recognition to diseases and disorders like Sleep Apnea, Epilepsy, and Alzheimer's disease.

Although various classification techniques have been reported in the literature Alhagry et al. [7] reports a classification accuracy of 85.65%, 85.45%, and 87.99% for valance, arousal, and liking, by using LSTM model. Similarly, Li et al. [8] employed LSTM model achieving a mean accuracy of 76.6%. We compute the band power feature for each frequency band of the EEG signals and employ the machine learning methods, namely Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Decision Tree and Random Forest. Section II for Related Work presents a detailed description of the previous research work. In this work, we achieve a maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, outperforming the other classifiers.

In the rest of the paper, Section II describes the related work. Section III contains the proposed methodology, and Section IV includes the experimental results. Finally, we conclude the paper with section V with the conclusion and future work.

II. Related Work

There are various emotions like happy, excited, angry, afraid, sad, depressed, calm, and contentment and the proper classification of these emotions can be beneficial for the study. Bastos-Filho et al. [9] have classified the emotional state as calm when the levels of arousal are below 4, and the level of valence is between 4 and 6. Similarly, for stress, the levels of arousal should be greater than 5, and that of valence should be less than 3.

In researches concerning the problem of emotion classification, the ability to classify emotions depends on two main factors:

1) Features extracted from the dataset.
2) Classifiers used for emotion classification.

The classification accuracy compared to the original dataset can be improved by extracting a wide range of features from the dataset. There are mainly three types of features :

1) Time-domain features.
2) Frequency domain features.
3) Time-Frequency domain features.

Jenke et al. [10] have described several features and their relevance to EEG signals. Some of the features are statistical features like mean, standard deviation, power. Hjorth Features like activity, mobility, complexity. Frequency domain features include band power, higher-order spectra. The time-frequency domain features are Hilbert-Huang Spectrum (HHS) and Discrete Wavelet Transform (DWT). Recent researches have shown that frequency domain features are more useful in the analysis of EEG signals. A good number of papers have used PSD, or PSD-based features generated from EEG signal datasets and achieved good accuracy to solve problems in the domain of emotion recognition and classification. Raphael Vallat also mentions the use of PSD for a myriad of analyses [11]. This motivates us to use frequency-domain features for extracting information from the EEG signals and explore various classification techniques.

This paper examines different classification techniques for Emotion Recognition on the publicly available DEAP (Dataset for Emotional Analysis using Physiological Signals) dataset [12]. We have also found Recurrent Neural Networks (RNN) being used in recent years to address the problem of emotion recognition and classification effectively. We mention some of the prominent researches employing LSTM model and other classifiers using DEAP dataset. However, all of them achieve accuracy less than the proposed model in this paper.

Dabas et al. [13] proposed a 3-D emotional model (Valence, Arousal, and Dominance) for classifying emotions using the DEAP dataset. They used machine learning algorithms like SVM, Naive Bayes, and achieved an accuracy of 58.90% and 78.06%.

Liu et al. [14] has employed the DEAP dataset for classifying emotions and features like time-domain features (mean, power, standard deviation, higher-order crossings, fractal dimension, Hjorth feature), frequency domain features (power spectral density), time-frequency domain feature (discrete wavelet transform). Multi-electrode features (differential asymmetry and rational asymmetry, magnitude squared coherence estimate) are computed and uses maximum relevance minimum redundancy (mRMR) for feature selection. KNN and RF are employed as classification techniques with the highest accuracy of 66.17% for arousal using a magnitude squared coherence estimate as a feature.

Wichakam et al. [15] have experimented on the DEAP dataset using band power as the feature and SVM as classifier. The maximum accuracy achieved is 64.9% for valence and 66.8% for liking while using the 3-dimensional emotion model. They have only used ten channels and have shown that performance accuracy is not improved even if 32 channels are employed.

Salama et al. [16] designed a 3-dimensional convolutional neural network for emotion recognition from multi-channel EEG data. They have used the DEAP dataset for analysis and have achieved 87.44% and 88.49% accuracy for valence and arousal classes.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:03:17 UTC from IEEE Xplore. Restrictions apply.

Alhagry et al. [7] have proposed an end-to-end model employing the LSTM classifier for emotion classification on the DEAP dataset. The average subject-independent accuracy achieved for arousal, valence, and liking is 85.65%, 85.45%, and 87.99%.

Li et al. [8] have used RASM as a feature which represents frequency-space domain characteristics of the EEG signal. They have employed the LSTM model on DEAP dataset and achieved a mean accuracy of 76.6%. Xing et al. [17] built Stack Autoencoder (SAE) for EEG signal decomposition and used LSTM model for classification but still observed accuracy of 81.1% in valence and 74.38% in arousal.

Yang et al. [18] propose a parallel combination of Convolutional Neural Network and LSTM Network to extract features from the DEAP dataset and then use the softmax classifier for classification. This model obtained the mean accuracy of 90.80% and 91.03% for valence and arousal.

Previous works in the Emotion Recognition establish that the LSTM models perform better than other classification techniques. Still, the reported work examines either the raw EEG signals or the time-domain features for training the LSTM model. Here, we investigate the use of the band power, a frequency-domain feature for training the LSTM model. To compare the performance of the proposed LSTM model with other classifiers, we also train classifiers, namely SVM, KNN, Decision Tree, and Random Forest. On contrasting the results, we observe a maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, which is significantly better than other classifiers.

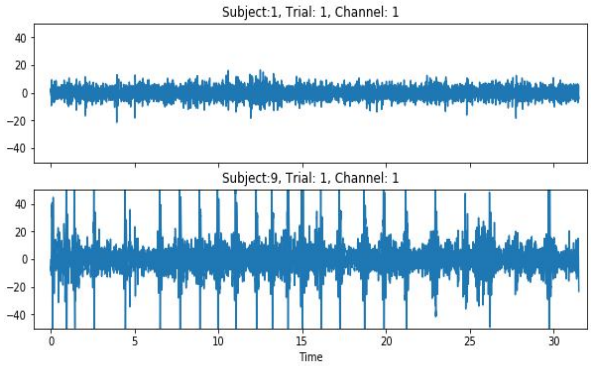


Fig. 1. EEG signal for Subjects:(1, 9), Trial: 1, and Channel: 1

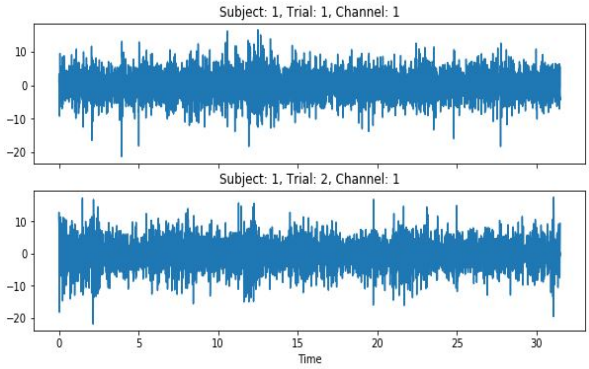


Fig. 2. EEG signal for Subject:1, Trials: (1, 2), and Channel: 1

III. Methodology

A. Dataset Description

The DEAP dataset [12] is a multimodal dataset for the determination of human emotional states available for public access. For the creation of the database, experiments were performed where 32 participants were made to watch 40 one-minute-long excerpts of selected videos, during which the signals from 32 channels of standard EEG headset and physiological signals from 8 channels were captured. Equipment had a sampling frequency of 128 Hz. The data was pre-processed, and artefacts were removed. The participants rated each video on a scale of 0-9 in terms of valence, arousal, dominance, and liking. These ratings become the benchmark for the classification of emotional states.

B. Pre-processing

Fig. 1 illustrates the EEG signals for two subjects (1, and 9) subjected to the same trial. Both signals indicate differences in the magnitude and the pattern of activations of the brain for different individuals. Thus, it displays the uniqueness in the processing of information in the brain for every individual.

Fig. 2 illustrates the EEG signals for a single person, subjected to two different trials. Both signals display similarity in the magnitude of the activation of the brain for the subject.

As inferred from Fig. 1 and Fig. 2, we conclude that emotion recognition is to be done for each individual separately as the analysis displays no similarity in the activation of the signals recorded for different subjects in the population. Since every individual possesses unique consciousness and emotional limits, the prediction of emotion for an individual using the learning from any other individual will drastically reduce not only the accuracy in prediction but the model will also lose its credibility for the prediction of an unknown subject. But the analysis also shows that there does exist similarity amongst the signals, valence and arousal values for the different trials of an individual. Hence, there is a possibility of finding a pattern for a certain emotion by understanding the signals obtained for that individual only. We exploit the results of this analysis to design a customized model for emotion recognition. Although by increasing the size of data, it is possible to account for this diversity in the strengths of the EEG signals. This is demonstrated by the modern extensive Image Classification datasets such as Tencent [19] which consists of more than 17 million images. Such a large dataset enables the Deep Learning models to explore the hidden features of the dataset and account for the diversity in the sample population. But the DEAP dataset consists of only 32 subjects, so it is currently not possible to account for such variance in the EEG signals. Therefore this study limits its training and testing to independent subjects.

The original signals were recorded for 63s (3s prior and 60s for the video). The preceding signal recorded is not removed as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:03:17 UTC from IEEE Xplore. Restrictions apply.

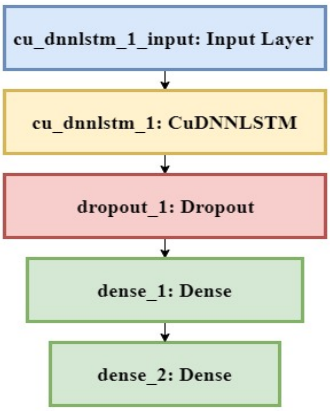


Fig. 3. Proposed LSTM Model

it does possess useful information regarding the state of mind of the individual before showing video trials. The band power for the different bands is calculated using the Welch method based on the Hanning window. The window length here is 1s and the stride of 0.25s for the entire 63s, therefore, obtaining 249 band power values at different instances of time. Only the EEG data values are used for experimental purposes, as our long term goal is to develop a real-time emotion prediction model where we require a minimal amount of hardware so that it can be used in the daily lives of every individual, especially patients.

C. LSTM Model

In this paper, we use the LSTM network as a useful tool for the prediction of the emotion of individuals. The LSTM networks are frequently used for handling sequential data such as paragraphs in NLP, previous electricity load in the case of the electricity demand prediction. LSTM cell possesses the ability to remember the distant as well as recent events to accurately predict the target variable. This property of retention can turn to be useful for emotion recognition as knowing about the past activations of the EEG signals can drastically affect the prediction of target variables and provide useful insights to the events leading to an appropriate response for the subject.

Fig. 3 shows the configuration of the Proposed LSTM model. We implement the model in Python 3 on the Google Colab platform with GPU support for the LSTM network. The LSTM layer has 40 nodes. Dense_1 layer has 10 nodes with 'tanh' activation function, and Dense_2 has a single node with 'sigmoid' activation function. We use a Dropout 25% between the LSTM Layer and Dense_1 Layer. We use Stochastic Gradient Descent (SGD) optimizer (learning rate=0.01, learning rate decay constant=$1 * 10^{-5}$, and momentum constant=0.9) to minimize the binary-cross-entropy loss function.

Other than the customized LSTM model, we also test the classification on the dataset using KNN, SVM, Decision Tree, and Random Forest. We evaluate the performance of each classifier on the pre-processed dataset. As expected from the previous studies, the proposed LSTM model outperforms the other classifiers by a huge margin.

IV. Results

To verify the effectiveness of the proposed LSTM model, we contrast the performance of KNN, SVM, Decision Tree, Random Forest, and LSTM models for classification of the preprocessed dataset. We test models several times to ensure the significance of the results observed. Table I highlights the average prediction accuracies.

TABLE I
Testing Accuracies

Model	Valence	Arousal
KNN	79.69	75.78
SVM	76.56	72.66
Decision Tree	77.34	74.21
Random Forest	80.46	77.34
LSTM	**94.69**	**93.13**

On analyzing the results, we observe that the LSTM model outperforms the other classifiers by a large margin. We observe a remarkable increment of about 16% and 18% for valence and arousal when comparing the LSTM model with other classifiers. The highest increment in average testing accuracy of 18% for valence and 20% for arousal is observed when comparing the SVM classifier with the LSTM model.

We compare our results with the results of Yang et al. [18] following similar experimental procedures with the parallel Convolutional Recurrent Neural Network model. Here, we observe an increment of 4% for valence and 2% for arousal in average testing accuracies for all the subjects. Our proposed model notes a significant increment of 9% in valence and 7.5% in arousal for Alhagry et al. [7] and 14% in valence and 19% in arousal for Xing et al. [17]. Fig. 4 illustrates the average testing accuracy for valence and arousal of the 32 subjects.

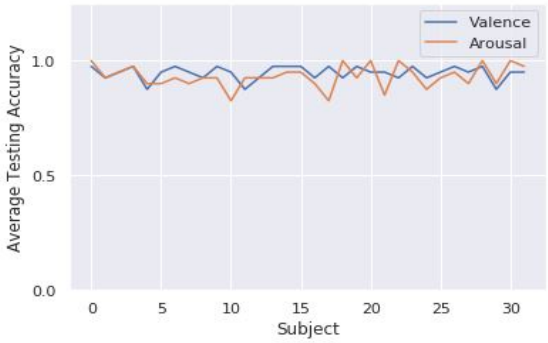


Fig. 4. Average testing accuracy of 32 subjects for LSTM model

V. Conclusion and Future Work

In this work, we evaluate power spectral density over the 32 channels of the DEAP dataset. We segregate them into

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:03:17 UTC from IEEE Xplore. Restrictions apply.

five bands of frequencies, namely Alpha, Beta, Gamma, Delta, and Theta, to derive the band power of each band. We use band power as a feature for classifying valence and arousal of the subject. We evaluate and compare using KNN, SVM, Decision Tree, Random Forest, and LSTM as our classifiers. On analysis, we observe a minimum average increment of 16% in the average testing accuracies and maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, which performs better than the current state-of-the-art classifiers.

In future, we can use the proposed experimental setup to obtain useful information regarding the emotions of the subjects and extend it for real-time applications. For further improvements, we can add more frequency domain features and test for their performance. As demonstrated by Wichakam and Vateekul [15], a subset of the channels for feature generation may perform better in terms of accuracy. The work may further be extended to include subject-independent models as well. The study can be also be extended to developing 3-D emotion models like the work of Dabas et al. [13].

REFERENCES

[1] R. Plutchik, *The Emotions*. University Press of America, 1991.

[2] J. A. Russell, "A circumplex model of affect," *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[3] M. M. Bradley, M. K. Greenwald, M. C. Petry, and P. J. Lang, "Remembering pictures: Pleasure and arousal in memory," *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 18, no. 2, pp. 379–390, 1992.

[4] W. David and T. Auke, "Toward a consensual structure of mood," *Psychological bulletin*, vol. 98, no. 2, pp. 219–235, 1985.

[5] K. Takahashi, "Remarks on SVM-based emotion recognition from multi-modal bio-potential signals," in *Proc. IEEE Int. Work. on Robot and Human Interactive Communication*, 2004, pp. 95–100.

[6] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals," *Multimedia Content Representation, Classification and Security. MRCS 2006. Lecture Notes in Computer Science*, vol. 4105, 2006.

[7] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[8] Z. Li, X. Tian, L. Shu, X. Xu, and B. Hu, "Emotion Recognition from EEG using RASM and LSTM," in *International Conference on Internet Multimedia Computing and Service*. Springer, 2017, pp. 310–318.

[9] T. F. Bastos-Filho *et al.*, "Evaluation of feature extraction techniques in emotional state recognition," in *Proc. IEEE Int. Conf. Intelligent human computer interaction*. IEEE, 2012, pp. 1–6.

[10] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," *IEEE Trans. on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[11] R. Vallat. Bandpower of an EEG signal. [Online]. Available: https://raphaelvallat.com/bandpower.html (accessed Nov. 13, 2019)

[12] K. Sander *et al.*, "Deap: A Database for Emotion Analysis; Using Physiological Signals," *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[13] H. Dabas, C. Sethi, C. Dua, M. Dalawat, and D. Sethia, "Emotion Classification Using EEG Signals," in *Proc. ACM Int. Conf. Computer Science and Artificial Intelligence*. ACM, 2018, pp. 380–384.

[14] J. Liu, H. Meng, A. Nandi, and M. Li, "Emotion detection from EEG recordings," in *Proc. IEEE Int. Conf. Natural Computation, Fuzzy Systems and Knowledge Discovery*. IEEE, 2016, pp. 1722–1727.

[15] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in EEG-based emotion prediction with support vector machines," in *Proc. IEEE Int. Conf. Joint conference on computer science and software engineering*. IEEE, 2014, pp. 106–110.

[16] E. S. Salama, R. A.El-Khoribi, M. E.Shoman, and M. A. Shalaby, "Eeg-Based Emotion Recognition using 3D Convolutional Neural Networks," *Int. Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018.

[17] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG," *Frontiers in Neurorobotics*, vol. 13, p. 37, 2019.

[18] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network," *Proc. IEEE Int. Conf. Joint Conference on Neural Networks*, 2018.

[19] B. Wu, W. Chen, Y. Fan, Y. Zhang, J. Hou, J. Huang, and T. Zhang, "Tencent ML-Images: A Large-Scale Multi-Label Image Database for Visual Representation Learning," *arXiv preprint arXiv:1901.01703*, 2019.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:03:17 UTC from IEEE Xplore. Restrictions apply.



Research Article

An experimental analysis of tensile, hardness and wear properties of aluminium metal matrix composite through stir casting process

Anubhav Coyal[1] · N. Yuvaraj[1] · Ravi Butola[1] · Lakshay Tyagi[1]

Received: 23 January 2020 / Accepted: 30 March 2020 / Published online: 13 April 2020
© Springer Nature Switzerland AG 2020

Abstract
Among several methods to fabricate Aluminium metal matrix composites, stir-casting technique is used for large-scale production due to its being less expensive. This research is concerned with the fabrication of Aluminium composites with SiC and Jute Ash Particles, for sake of producing a cost-effective composite with enhanced properties. Wear behaviour was studied using pin on disc tribometer and it was found that wear resistance increases with-addition of reinforcement particles. Fabricated composite samples showed almost 4 times better wear resistance compared to base material. Matrix metal reinforced with SiC showed the best wear resistance of all fabricated samples. Coefficient of friction significantly decreased with the presence of reinforcement due to formation of mechanical mixed layer. Tensile Strength was escalated with the introduction of reinforcement as load is transferred to strongly bonded reinforcement particles. Microhardness was enhanced with the introduction of reinforcement and Sample 2 with SiC as reinforcement showed the best microhardness. Microstructure and fractography of the base metal and fabricated composites was analyzed by Scanning Electron Microscope (SEM).

Keywords Composites · Wear · Tensile · Hardness

1 Introduction

Composite is a mixture of two or more materials which are not soluble in one another, they have superior properties compared to any of its component materials. Composite materials have acquired a large section of automotive industry due to their light weight & robust behavior i.e. being able to withstand harsh conditions [1]. Aluminium Metal Matrix Composites (AMMC) are used for various applications such as brake rotors, pistons, cylinder liner [2]. These are also used in aerospace and marine industries because of its lower production cost and balanced properties [3]. AMMC can be fabricated through various processes ranging from liquid metallurgy processing and solid metallurgy processing, some liquid metallurgy processing are compos casting, stir casting, liquid infiltration etc. [4]. Out of all liquid metallurgy processing ways out there, stir casting is one of the most flexible as it allows us react in case of change, less expensive and have the higher production rate [5].

Recent studies show that various natural fibre products can be utilized for advance engineering application. Mahesh et al. [6] fabricated jute rubber composite with different stacking sequences and studied the influence of process parameters on wear behavior of composite. They concluded that wear rate is most affected by abrading distance, compared to load and composite configuration. Praveen Kumar et al. [7], used Al-4.5%Cu alloy and reinforced it with 2%, 4% and 6% of bamboo leaf ash (BLA) respectively via stir casting. Slurry erosion behavior of hybrid matrix flexible composite (HyMFC) made from different stacking sequence of jute and rubber, and

✉ Ravi Butola, ravibutola33855@gmail.com | [1]Department of Mechanical, Production and Automobile Engineering, Delhi Technological University, Delhi 110042, India.

single matrix flexible composite (SMFC) were compared by Mahesh et al. Process parameter were optimized using Taguchi and they concluded that best erosion resistance in slurry medium was provided by HyMFC composed of vulcanized rubber bonding gum [8]. Ravi et al. [9], fabricated Aluminium metal matrix composite (AMMC) with natural fibre reinforcements like bagasse, GSA, RHA, coconut shell ash using stir casting.

Earlier many researchers have used stir casting to fabricate composites materials reinforced with various ceramic particle reinforcements, and found that fabricated composites have enhanced mechanical and wear performance [10–16]. Anshul et al. [17] reviewed the mechanical properties of variety of alloys like Mg–4Y–3Nd (WE43), Mg–ZrSiO4–N_2O_3, Al–Si hypoeutectic A356 alloy, 5210 steel (WC-12% CO coated) under the influence of friction stir processing. They concluded that in reinforced aluminium alloys wear resistance increased. Ravi et al. [18] studied different effects on mechanical and tribological properties of AMMC reinforced with SiC, boron halide, fly ash, Graphite and Rice husk ash. Kaushik et al. [19], studied the abrasive wear behavior of stir casted Al6082–SiC–Gr composites and compared it with Al 6082–SiC composites and Al6082 alloy. Wear resistance of Al–SiC–Gr hybrid composites surpassed all other fabricated composites. Alaneme et al. [20] studied the wear behavior Aluminium matrix hybrid composites reinforced with alumina, Rice husk ash and graphite. Composites without graphite showed better wear properties. Marigoudar et al. [21] studied the wear behavior of zin–Aluminium alloy reinforced with SiC. By integrating SiC particles wear resistance properties of the material were observed to enhance. It was also witnessed that with increase in SiC content the wear of composite decreased.

Md. Tanwir Alam et al. [22] fabricated AA356 composite using varying wt% of SiC nanoparticles as reinforcement via Stir Casting. They found that tribological properties were enhanced with increasing wt% of SiCn, and adhesive wear was the major wear mechanism. They also performed regression analysis and successfully developed a mahematical function for the same. Rao et al. [23] found that Al/SiC composite had significantly lesser wear coefficient than the alloy and it subdued further due to pressure applied and SiC particulate addition. Jo et al. [24] fabricated a hybrid metal matrix composite of aluminium–SiC–B4C, ballistic impact tests & various compressive test were performed. Hybrid composite was found to perform better in all the performed tests. Alam et al. [25] used two step stir casting to reinforce AA356 with SiCn, optimization and analysis of controlling factors were done using Taguchi and ANN. By SEM analysis they found that wear was decreased due to the presence of oxide layer and SiCn particles. It was concluded that Taguchi and ANN can be used to decrease time involvement in industries in various testing processes.

In order to fulfill the needs of the growing population there is an increase in agricultural activities resulting in increasing agricultural waste production. This agricultural waste needs to be disposed off, so in this research an attempt has been made to utilize this freely available natural fibre ash utilize for engineering application. In this work, tribological and mechanical properties of aluminium metal matrix composite reinforced with ceramic and natural fibre (Jute ash) has been studied. Jute is a lignocellulose fibre having promising mechanical and thermal properties. Its light-weight, less cost and ease of availability made it ideal to use it as reinforcement, using these type of natural fib re material as reinforcement is eco-friendly way of dealing the natural fibre [26–28], whereas SiC are ceramic particles used in order to increase the hardness of the composite.

2 Materials and method

2.1 Materials

In this study, the material selected is AA6061 was taken as the matrix material, chemical composition is mentioned in Table 1 [13]. Reinforcements used in this study are SiC having a particle size of 10 µm shown in Fig. 1a. Jute ash is prepared with the exterior part of the coconut known as jute shown in Fig. 1b. The jute was first made dried in the sun rays for 7 days. Dried part is burnt using Camphor balls and jute ash is prepared of fine size of 50 µm, by ball milling as shown in Fig. 1c.

2.2 Fabrication of composite

The stir casting experimental setup mainly comprised of an electric furnace and stirrer. The furnace can work with working temperature of 1000 °C and the working current rating of furnace is single phase 230V AC, 50 Hz. The AA6061 is taken in the form of fine scraps then, the scrap is placed in the furnace and heated to a temperature just above 600 °C and converted to semi-liquid state. The

Table 1 AA6061 chemical composition

Elements	Si	Fe	Cu	Mn	Mg	Zn	Cr	Ti	Al
wt%	0.64	0.29	0.26	0.1	0.88	0.03	0.09	0.03	Rest

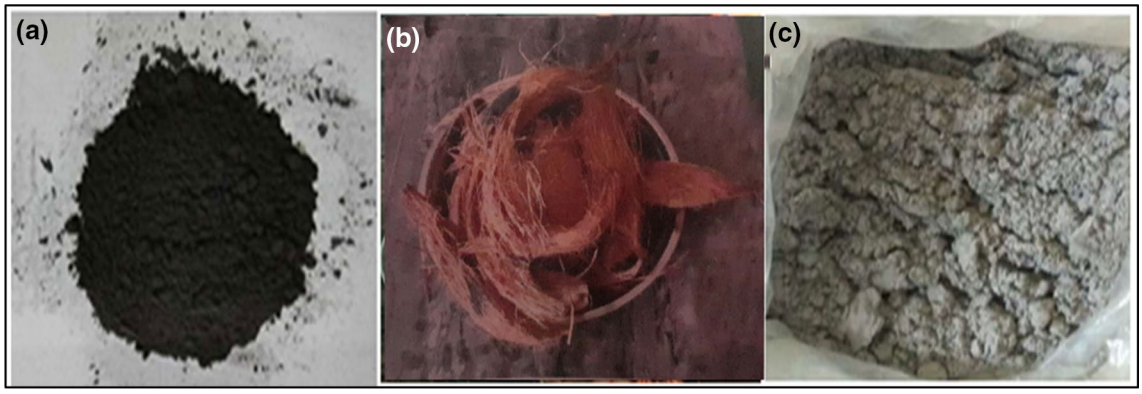


Fig. 1 **a** SiC powder **b** Dried Jute **c** Jute Ash after burning and ball milling

Table 2 Sample composition

Samples	Al alloy 6061%	Silicon carbide (%)	Jute Ash (%)
Sample 1	100	0	0
Sample 2	96.6	3.4	0
Sample 3	99.1	0	0.9
Sample 4	95.7	3.4	0.9

process is followed by the addition of preheated reinforcement powder which is heated to a temperature of 5000 °C, to the semi liquid matrix, and argon gas is made to enter into the furnace for 3 min. Mechanical stirring is used at a rotating speed of 150 rpm while reheating process is going on and this process is carried out for 14 mins. Then the molten AMMC is poured into the cylindrical shape molds and subjected to compaction to produce the required Sample. Four Samples were made, first composed of just the base material (AA6061), second consist 3.4 wt% of SiC taken on basis of the literature review, third Sample consisted of randomly take < 1 wt% of jute ash and forth Sample consisted of both jute ash and SiC, composition of Samples is mentioned in Table 2. The Schematic representation of the stir casting setup is shown in Fig. 2.

2.3 Tensile testing

Sample for tensile were carved through the Wire EDM process. Large bone shaped tensile samples were drawn up according to ASTM: E8/E8M-011, and the prepared tensile sample are shown in Fig. 3. UTS and percentage elongation was given as the output by Tinius Olsen H50KS, computer controlled universal testing machine which was used to examine the tensile behavior of the samples.

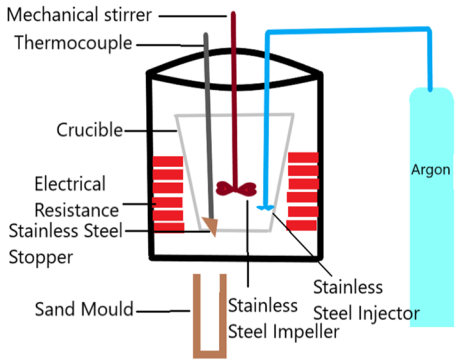


Fig. 2 Schematic representation of the stir casting setup

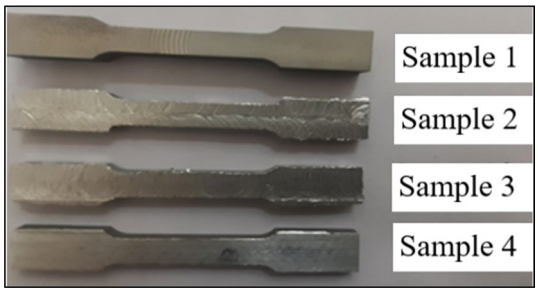


Fig. 3 Tensile Samples

2.4 Microhardness

Fisher scope HM2000S was used for studying microhardness of the prepared Samples, microhardness can be tested for materials ranging from very soft to hard, and testing can also be done on heavy loads. Microhardness was studied at 4 different points for constant 300 gm load for 20 s according to ASTM E8/E8M-011 standard. The average of all 4 values for a single Sample was taken as the resulting microhardness.

2.5 Wear testing

Pins of diameter 10 mm were carved out from the casted composite, in order to test the wear on pin on disc tribometer (DUCOM) as shown in Fig. 4. Disc made of material EN 24 of diameter 100 mm, hardness 58 HRC and roughness 0.2 µm was used, wear test was conducted at load of 20 N on two different track diameter and rpm mentioned in Table 3, with a sliding speed of 2 m/s and sliding distance of 3000 m, this test was conducted according to ASTM G99-04 standard. These test conditions were taken as reference from the previous studies done by the researchers, all test conditions are mentioned in Table 3.

Table 3 Wear test conditions

S. no	Test conditions	Values
1	Load	20 N
2	Pin diameter	10 mm
3	Track diameter	80, 60 mm
4	Sliding lance	3000 m
5	rpm	238, 318
6	Disc roughness	0.2 µm
7	Disc hardness	58 HRC

3 Results and discussion

3.1 Tensile strength

Stress Strain relation and tensile test results of the Samples can be seen in Fig. 5. It can be concluded that total elongation of the fabricated composites is more than that of base metal, it means ductility has improved with the introduction of reinforcements. This change in ductility can be attributed to proper dispersion and abutting packing of reinforcement particle in the matrix [29]. There is drastic increase in tensile strength of the fabricated samples compared to the base material. This can be attributed to transfer of load to strongly bonded reinforcement particles in matrix [30]. Sample 2, having SiC as reinforcement showed the best results with an Ultimate Tensile Strength (UTS) of 123 MPa. Similar types of results were reported from kumara et al. [29], aluminium metal matrix composite materials reinforced with SiC. Whereas, Sample 3, with Jute Ash as reinforcement showed a UTS of 122MPa followed by sample 4, having both SiC and Jute Ash as reinforcement with UTS of 113 MPa. Adhesion of matrix and reinforcement effect mechanical properties [31], the above decrease in UTS might be due to the improper adherence when jute ash and SiC are reinforced simultaneously with the matrix metal. When reinforced jute epoxy reinforced composite with high wt% of Al2O3 found tensile strength to decrease because of improper adhesion of filler and matrix [32]. Toughness of a sample is corresponding to its UTS values i.e. sample having more toughness will have higher UTS value. Sample 2 is toughest of all fabricated samples followed by sample 3, then sample 4 and at last sample 1 or base material has the minimum toughness.



Fig. 4 Pin on disc tribometer

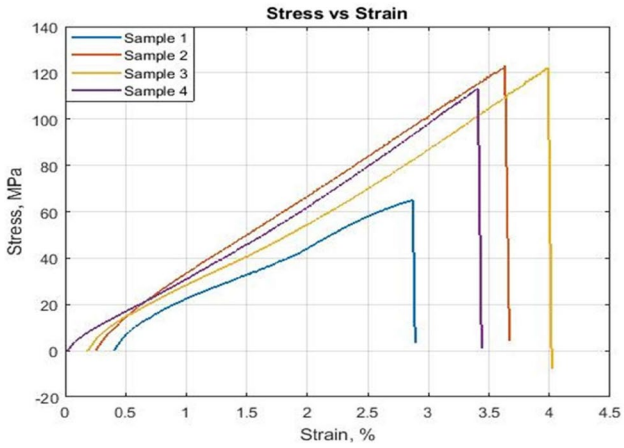


Fig. 5 Stress–Strain diagram of all the samples

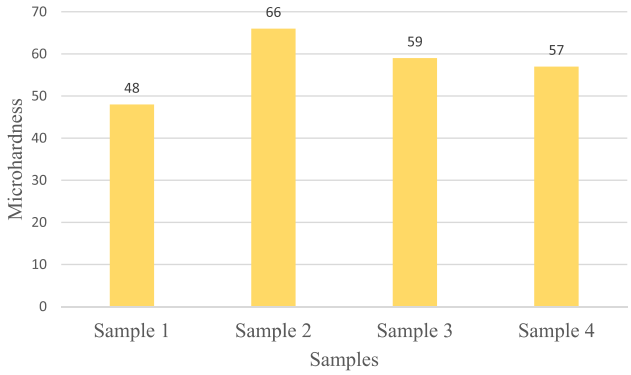


Fig. 6 Microhardness of Samples

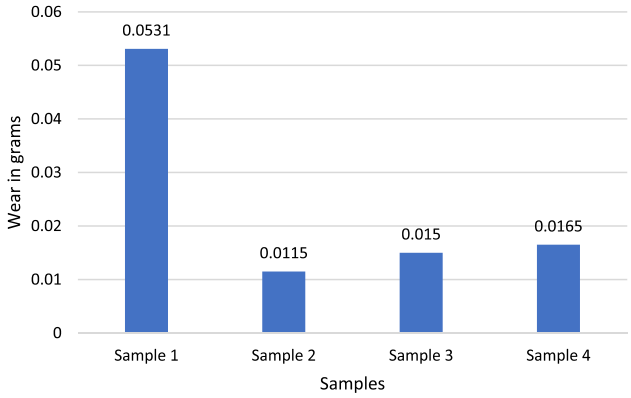


Fig. 7 Wear loss in grams

3.2 Microhardness

From Fig. 6 it can be observed that with the introduction of reinforcement microhardness of the base material was enhanced. Sample 2, having only ceramic as reinforcement showed the best microhardness of 66 HV as ceramic are hard. These ceramic particles hinder the movement of dislocation hence increasing the microhardness. This increase in microhardness might also be ascribed to dynamic process of recovery, re-crystallization [29, 30]. Sample 3 had better microhardness than base material because of presence of hardness of ash particles. Similar type of results were reported by Butola et al., when reinforced Al 6061 with various natural fibres reinforcement [5]. The recorded microhardness of Sample 3 was less than that of Sample 2 as jute ash is not as hard particles as compared to SiC particles. Sample 4 had the least microhardness than all other fabricated composite samples. This degradation of mechanical property might be due to the unseemly adhesion of jute ash and SiC particles with the matrix [31].

3.3 Wear

Wear loss of the tested samples is shown in Fig. 7, with the introduction of reinforcement wear loss decreased. Minimum wear loss was recorded for Sample 2 which contains only SiC as reinforcement; as well dispersed SiC particles bear most of the load and allow minimal metal to metal contact [23]. Sample 3 showed less wear loss than other samples but more than the Sample 2, as scattered jute ash bears most of the load but these are not as hard as compared to SiC particles. Sample 4 had the highest wear loss of all composite samples because of increased ductility of the Sample, still its wear loss was almost 70% less than that of the base material. Enhanced wear resistance of the fabricated samples is corresponding to their increased hardness. Similar type of results were reported by Mahesh et al. [33] assimilated areca nut filler in coir fiber reinforced areca nut and reported that main factor influencing the wear resistance of any composite is its hardness and results.

3.4 Coefficient of friction as a function of sliding distance

The friction coefficient significantly decreased with the presence of reinforcement, sliding wear counterparts leads to the formation of carbide particles layer on the surface and friction coefficient is decreased. A tribo-layer i.e. mechanical mixed layer (MML) is formed which reduces the direct contact surface between the disc and the composite as a result friction coefficient and wear loss is decreased. The variation of friction coefficient with the sliding distance for all tested wear is shown in Fig. 8. Initially there are big spikes, due to the initial metal to metal contact after the wear of base material when reinforcement comes to the surface due to the introduction of tribo-layer steadiness comes in the plot.

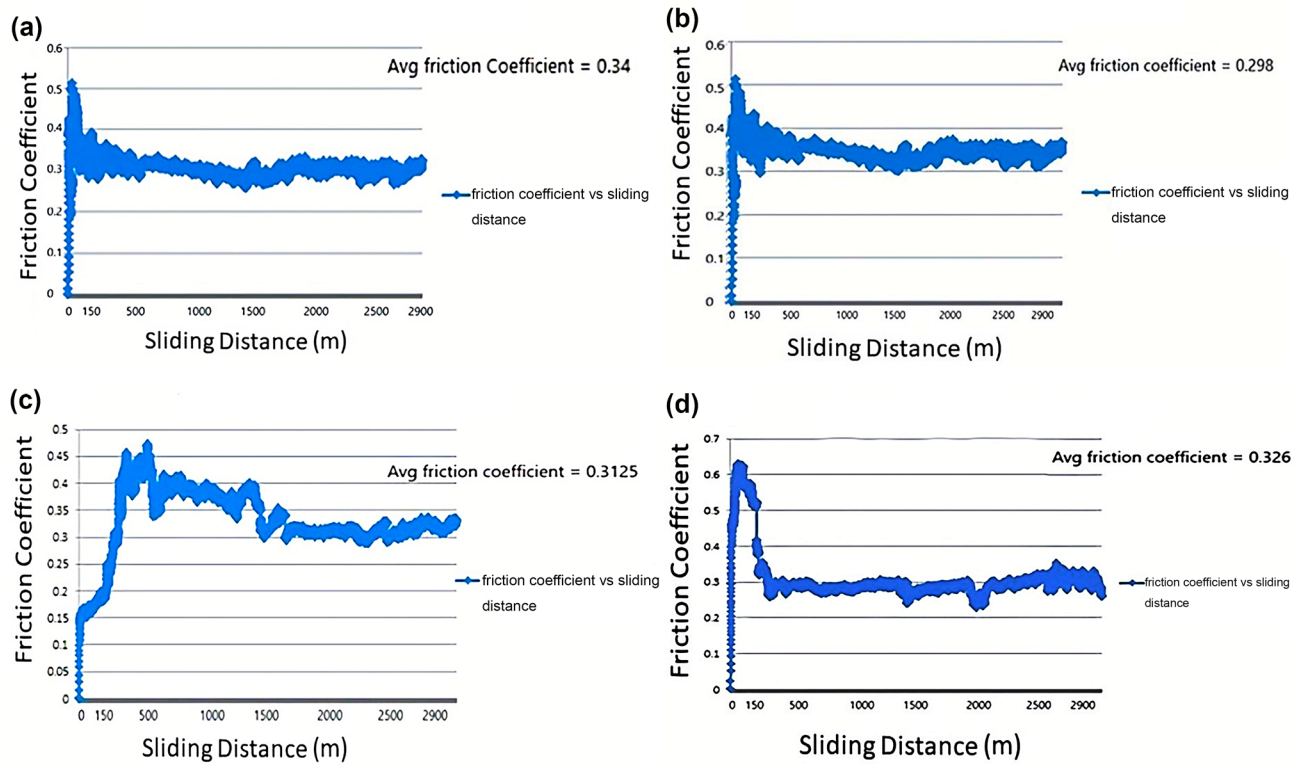


Fig. 8 Coefficient of friction v/s sliding distance of **a** Sample 1, **b** Sample 2, **c** Sample 3, **d** Sample 4

3.5 Microstructure analysis

Scanning Electron Microscope (SEM) was done in order to get high quality microstructure image of the surface of the base metal and surface of the fabricated composite samples. In order to get more refined grain boundaries on the surface, the surface was polished with different grade emery paper ranging from lower to higher grade i.e. 200, 400, 600, 800, 1000, 1200, 1600 and 2000. After this wet polishing was done in order to get fine finished surface, and then SEM images were generated. A SEM image confirms the existence of SiC and jute ash particles in the matrix and at some places clustering of particles can be seen. Grain boundaries can be seen in the microstructure images in Fig. 9a–d of the Samples. Increase in wear resistance and hardness can be attributed to distribution of reinforcement particles in the matrix which is confirmed in the SEM images, similar type of results were reported by other researchers [22].

3.6 Fractography analysis

After tensile test fractured surfaces were examined using scanning electron microscope and fractography images obtained are shown in Fig. 10. It can be asserted that the main reason of fracture was the presence of voids and tearing of matrix. In Fig. 10b crack formation on the fractured surface can be observed. This might be attributed to

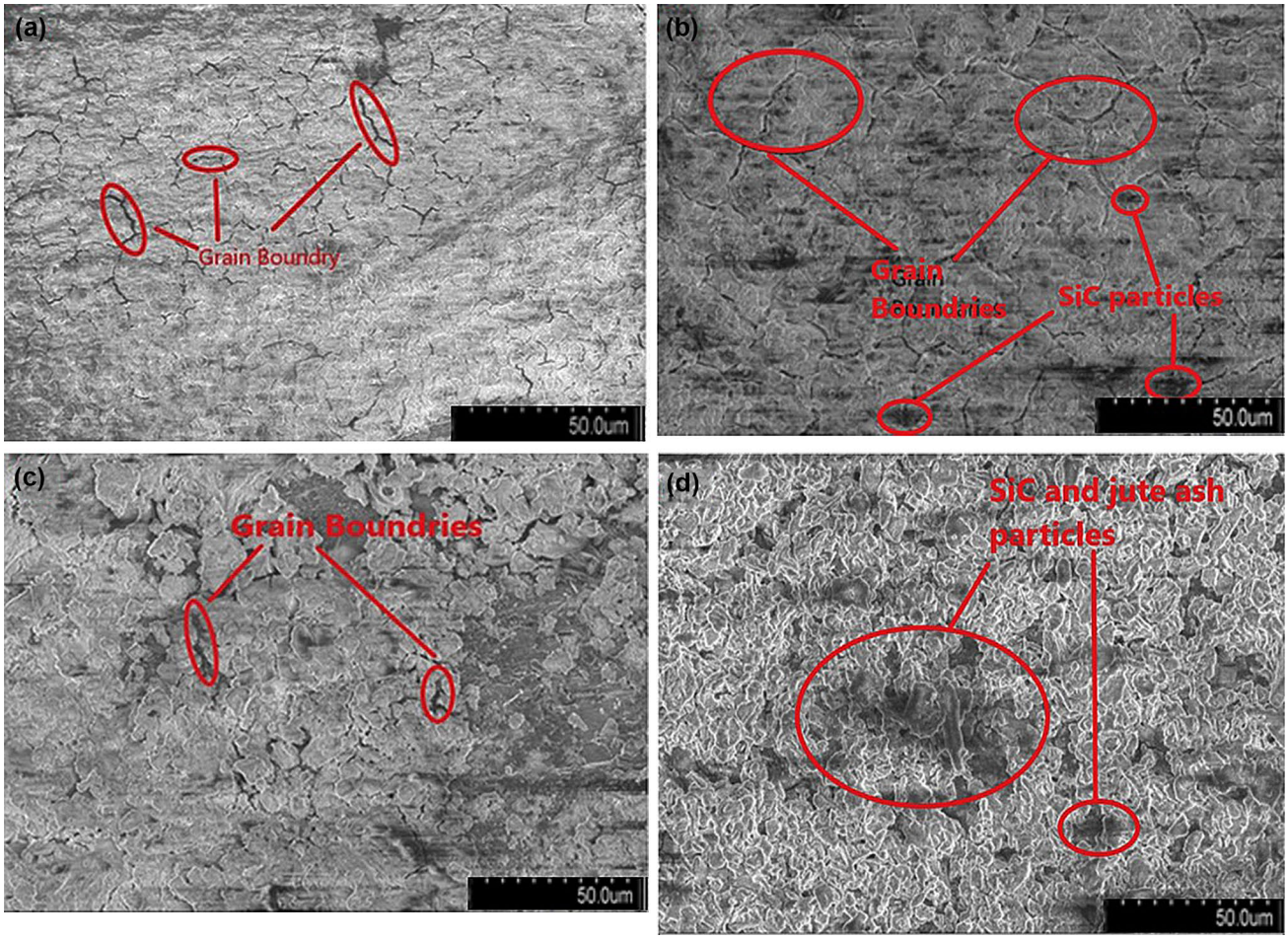


Fig. 9 SEM images of **a** Sample 1, **b** Sample 2, **c** Sampe 3, **d** Sample 4

the resistance offered by the hard SiC particles dispersed all along the matrix. Small dimples on the fractured surface can be seen in Fig. 10c. These dimples indicate that the composites have experienced the ductile fracture, enhanced ductility has also been discussed in this study earlier. Craters formation can be observed in Fig. 10d. These might have formed due the lumping of the reinforcement, this clustering can also be observed in Fig. 9d. Similar pores have been reported by K. Logesh et al. [34] in their study. These craters support and accelerate the crack formation resulting in untimely fracture.

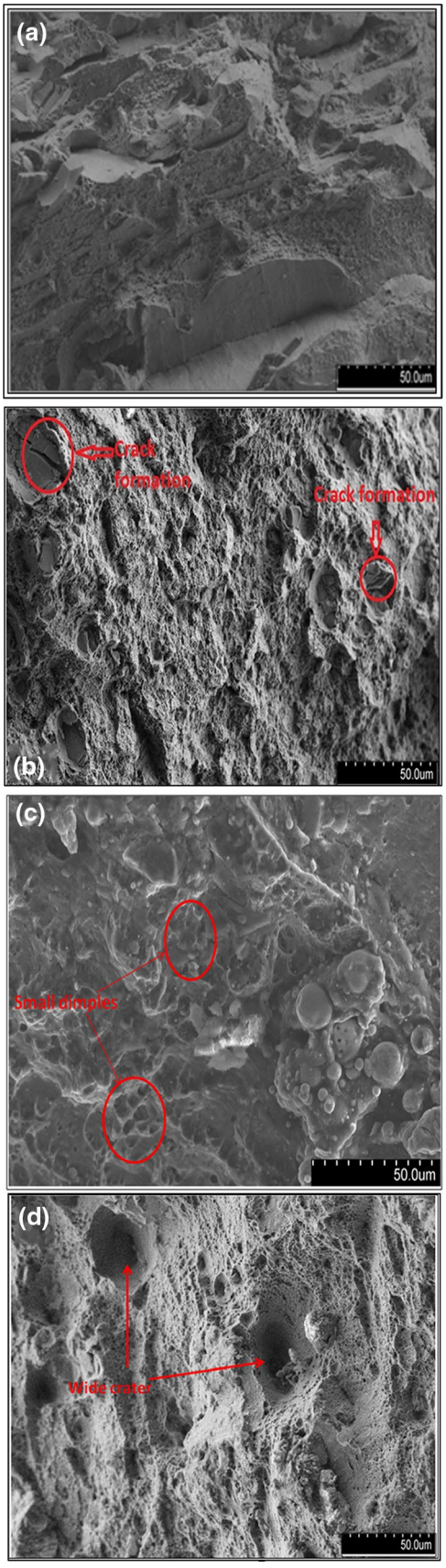


Fig. 10 Fractography images of **a** sample 1, **b** Sample 2, **c** Sample 3, **d** Sample 4

4 Conclusion

In this research, attempts have been made to fabricate a composite using jute ash and SiC as reinforcement. The fabricated composites showed enhanced tensile and tribological properties than its matrix material. The following conclusion can be drawn from this study:

- Tensile Strength and elongation increased with the introduction of reinforcement due to proper dispersion of reinforcement particles.
- Out of all fabricated composite samples, sample 2 showed the maximum UTS value as the load is transferred to strongly bonded reinforcement, releasing most of the stress from the Al matrix.
- Sample 2 showed the minimum wear loss as SiC beared most of the load, followed by sample 3, and then sample 4, which had significantly less wear loss than base material.
- Wear resistance increased with introduction of reinforcement as wear is dependent on hardness and reinforcements used are harder than the matrix.
- Microhardness was improved with the presence of reinforcement particles and all reinforced samples had microhardness more than the base material.
- Sample 2 showed the maximum microhardness because of presence of hard SiC. Sample 3 showed the minimum microhardness of all reinforced samples, because of improper adhesion of reinforcements with the matrix.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Safri S, Sultan M, Jawaid M, Jayakrishna K (2018) Impact behaviour of hybrid composites for structural applications: A review. Compos B Eng 133:112–121. https://doi.org/10.1016/j.compositesb.2017.09.008
2. Gurusamy P, Prabu S, Paskaramoorthy R (2015) Influence of processing temperatures on mechanical properties and microstructure of squeeze cast aluminum alloy composites. Mater Manuf Proces 30(3):367–373. https://doi.org/10.1080/10426914.2014.973587
3. Yuvaraj N, Aravindan S (2016) Wear characteristics of Al5083 Surface hybrid nano-composites by friction stir processing. Trans Indian Inst Met 70(4):1111–1129. https://doi.org/10.1007/s12666-016-0905-9
4. Butola R, MS Ranganath Murtaza, Q., (2019) Fabrication and optimization of AA7075 matrix surface composites using Taguchi technique via friction stir processing (FSP). Eng Res Express. https://doi.org/10.1088/2631-8695/ab4b00
5. Butola R, Pratap C, Shukla A, Walia R (2019) Effect on the mechanical properties of aluminum-based hybrid metal matrix composite using stir casting method. Mater Sci Forum 969:253–259. https://doi.org/10.4028/www.scientific.net/msf.969.253
6. Mahesh V, Joladarashi S, Kulkarni M, S. (2019) Experimental study on abrasive wear behaviour of flexible green composite intended to be used as protective cladding for structures. Int J Modern Manuf Technol 11(1):1–2
7. Kumar B, Birru A (2017) Microstructure and mechanical properties of aluminium metal matrix composites with addition of bamboo leaf ash by stir casting method. Trans Nonferrous Metals Soc China 27(12):2555–2572. https://doi.org/10.1016/s1003-6326(17)60284-x
8. Mahesh V, Joladarashi S, Kulkarni S (2020) Slurry erosive study and optimization of material and process parameters of single and hybrid matrix flexible composites using Taguchi approach. In: Proceedings of The 35th International Conference of the Polymer Processing Society
9. Butola R, Malhotra A, Yadav M, Singari R, Murtaza Q, Chandra P (2019). Experimental Studies on Mechanical Properties of Metal Matrix Composites Reinforced with Natural Fibres Ashes. SAE Technical Paper 2019-01-1123. https://doi.org/10.4271/2019-01-1123
10. Poovazhagan L, Jayakumar K, Bharat R, Viswanathan K, Akshay M, Jaikumar A (2016) Synthesis and machining characterization of ultrasonication assisted stir cast SiCp reinforced aluminum nanocomposites. Mater Today 3(6):2339–2346. https://doi.org/10.1016/j.matpr.2016.04.145
11. Pazhouhanfar Y, Eghbali B (2018) Microstructural characterization and mechanical properties of TiB2 reinforced Al6061 matrix composites produced using stir casting process. Mater Sci Eng, A 710:172–180. https://doi.org/10.1016/j.msea.2017.10.087
12. Ravi B, Naik B, Prakash J (2015) Characterization of aluminium matrix composites (AA6061/B4C) fabricated by stir casting technique. Mater Today 2(4–5):2984–2990. https://doi.org/10.1016/j.matpr.2015.07.282
13. Ezatpour H, Sajjadi S, Sabzevar M, Huang Y (2014) Investigation of microstructure and mechanical properties of Al6061-nanocomposite fabricated by stir casting. Mater Des 55:921–928. https://doi.org/10.1016/j.matdes.2013.10.060
14. Umanath K, Selvamani S, PalanikumarK K, Niranjanavarma D (2014) Metal to metal worn surface of AA6061 Hybrid Composites Casted By Stir Casting Method. Procedia Engg 97:703–712. https://doi.org/10.1016/j.proeng.2014.12.300
15. Kandpal B, Kumar J, Singh H (2017) fabrication and characterisation of al2o3/aluminium alloy 6061 Composites fabricated by Stir casting. Mater Today 4(2):2783–2792. https://doi.org/10.1016/j.matpr.2017.02.157
16. Uthayakumar M, Aravindan S, Rajkumar K (2013) Wear performance of Al–SiC–B4C hybrid composites under dry sliding conditions. Mater Des 47:456–464. https://doi.org/10.1016/j.matdes.2012.11.059
17. Chaudhary A, Kumar Dev A, Goel A, Butola R, Ranganath M (2018) The Mechanical properties of different alloys in friction stir processing: a review. Mater Today 5(2):5553–5562. https://doi.org/10.1016/j.matpr.2017.12.146
18. Butola R, Singari RM, Bandhu A, Walia RS (2017) Characteristics and Properties of different reinforcements in hybrid aluminium composites: a review. Int J Adv Product Ind Eng-SI-MM 511:71–80
19. Kaushik N, Rao R (2016) Effect of grit size on two body abrasive wear of Al 6082 hybrid composites produced by stir casting method. Tribol Int 102:52–60. https://doi.org/10.1016/j.triboint.2016.05.015
20. Alaneme K, Sanusi K (2015) Microstructural characteristics, mechanical and wear behaviour of aluminium matrix hybrid composites reinforced with alumina, rice husk ash and graphite. Eng Sci Technol Int J 18(3):416–422. https://doi.org/10.1016/j.jestch.2015.02.003
21. Marigoudar R, Sadashivappa K (2011) Dry sliding wear behaviour of SiC Particles reinforced zinc-aluminium (ZA43) alloy metal matrix composites. J Minerals Mater Charact Eng 10(05):419–425. https://doi.org/10.4236/jmmce.2011.105031
22. Alam M, Arif S, Ansari A (2018) Wear behaviour and morphology of stir cast aluminium/SiC nanocomposites. Mater Res Express. https://doi.org/10.1088/2053-1591/aab7b3
23. Rao R, Das S (2010) Wear coefficient and reliability of sliding wear test procedure for high strength aluminium alloy and composite. Mater Des 31(7):3227–3233. https://doi.org/10.1016/j.matdes.2010.02.017
24. Jo M, Choi J, Yoo J, Lee D, Shin S, Jo I, Lee S, Lee S (2019) Novel dynamic compressive and ballistic properties in 7075–T6 Al-matrix hybrid composite reinforced with SiC and B4C particulates. Compos B Eng https://doi.org/10.1016/j.compositesb.2019.107041
25. Alam M, Arif S, Ansari A, Alam M (2019) Optimization of wear behaviour using Taguchi and ANN of fabricated aluminium matrix nanocomposites by two-step stir casting. Mater Res Express. https://doi.org/10.1088/2053-1591/ab0871
26. Mahesh V, Joladarashi S, Kulkarni S (2019) An experimental study on adhesion, flexibility, interlaminar shear strength, and damage mechanism of jute/rubber-based flexible "green" composite. J Thermoplast Compos Mater. https://doi.org/10.1177/0892705719882074
27. Mahesh V, Joladarashi S, Kulkarni S (2019) An experimental investigation on low-velocity impact response of novel jute/rubber flexible bio-composite. Compos Struct https://doi.org/10.1016/j.compstruct.2019.111190
28. Mahesh V, Joladarashi S, Kulkarni S (2019) Physio-mechanical and wear properties of novel jute reinforced natural rubber based flexible composite. Mater Res Express. https://doi.org/10.1088/2053-1591/ab0164
29. Mohanakumara K, Rajashekar H, Ghanaraja S, Ajitprasad S (2014) Development and mechanical properties of SiC reinforced Cast

and extruded Al based metal matrix composite. Procedia Mater Sci 5:934–943. https://doi.org/10.1016/j.mspro.2014.07.381
30. Rahman M, Rashed H (2014) Characterization of silicon carbide reinforced aluminum matrix composites. Procedia Eng 90:103–109. https://doi.org/10.1016/j.proeng.2014.11.821
31. Gwon J, Lee S, Chun S, Doh G, Kim J (2010) Effects of chemical treatments of hybrid fillers on the physical and thermal properties of wood plastic composites. Compos A Appl Sci Manuf 41(10):1491–1497. https://doi.org/10.1016/j.compositesa.2010.06.011
32. Ahmed K, Mallinatha V, Amith S (2011) Effect of ceramic fillers on mechanical properties of woven jute fabric reinforced epoxy composites. J Reinf Plast Compos 30(15):1315–1326. https://doi.org/10.1177/0731684411420606
33. Mahesh V (2019) Influence of areca nut nano filler on mechanical and tribological properties of coir fiber reinforced epoxy based polymer composite. Sci Iran. https://doi.org/10.24200/sci.2019.52083.2527
34. Logesh K, Hariharasakthisudhan P, Arul Marcel Moshi A, Rajan B, Sathickbasha K (2019) Mechanical properties and microstructure of A356 alloy reinforced AlN/MWCNT/graphite/Al composites fabricated by stir casting. Mater Res Express. https://doi.org/10.1088/2053-1591/ab587d

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Analysis of Congestion Control Mechanism for IOT

Aastha Maheshwari
Department of Computer Science & Engineering
Delhi Technological University, Delhi, India
aasthamaheshwari.am@gmail.com

Rajesh Kumar Yadav
Department of Computer Science & Engineering
Delhi Technological University, Delhi, India
rkyadav@dtu.ac.in

Abstract- In IoT (Internet of Things) network, a big amount of data is generated within a period of time.Hence it is required to critically consider and design a load balancing protocol. In this paper we survey different congestion control mechanisms designed for IoT based network, classifiedin two major categoriesi.e. protocol dependent and performing offloading. These classifications are based on technique used to balance load and avoid congestionrespectively. Protocol dependent approach is further classified as application layer protocol (CoAP) or network layer (RPL) protocol. These techniques improvise CoAP and RPL protocols to handle congestion issues. Offloading dependent approach covers different methods to balance the load evenly within a network. This analysis also includes the major concerns and the focus of different techniques to achieve congestion control within an IoT network.

Keywords—IoT Network, congestion control, offloading, CoAP, RPL, load balancing.

I INTRODUCTION

IoT seems to be the physical network that gadgets have been accessed via the Internet. Such things contain implanted creativity to work either with the inside states or the outside situation. It changes how and where choices are made and who makes them at the point when these things sense and communicate. This is an advance innovation in wireless communication with different space area for its application area[1][2]. The important thought of the whole concept is the prevalent existence across us about a variety of things or item – such as Radio-Frequency Identification (RFID) tag, actuator, cell phone, sensor, and so on that can connect with each other and work with their neighbors to achieve shared goals through one of a kind that tends to plan[3]. It refers to the constantly producing physical network thing that has a related IP address to internets accessibility and tendency, and the interaction that occurs between such item and other internet-enabled gadget and framework that empower applications in healthcare tracking, production line scheduling, and so on forth in some decision-making processes.

IoT supports numerous human life applications, making their world simpler, more secured and more fantastic. For eg, smart urban community, transport and homes are numerous applications. With the help of IoT, the urban area can be enhanced at numerous levels by enhancing the frameworks, improving public transport, reducing traffic congestion, and defending citizens who are healthy and gradually occupied with the network[4-6]. The advances made by Wireless-Fidelity (Wi-Fi) in home automation were used mainly because of the networked idea of sending electronics where, e.g., electronic gadgets, mobile gadgets, TV and so on are usually supported by Wi-Fi. Wireless-Fidelity has started to end up a few portions of the home Internet Protocol network and the growing pace of mobile computing gadgets such as tablets, smartphones, and so on[7-9]. In addition, an adjacent recognition necessary by IoT monitoring advances to hospitalized patients whose psychological status should be consistently observed can be performed on an ongoing basis. For smart health, sensors are used to collect extensive behavioral data and use cloud and gateway to investigate and storing data and then wirelessly send the broken information to parental figure for further inspection and survey [10-12]. The way that IoT applications and situations laid out above are intriguing which gives advancements to smart everything. , however, there are a few difficulties to the utilization of the IoTs idea in the expense of execution. IoT is additionally looked with numerous different difficulties, for example, Self-Organizing, Scalability, Data volumes, Power supply, Data interpretation, Wireless communications and Interoperability [13] [14]. In a certain case, a huge amount of IoT gadgets are depicted by low processing speed and low storage leading to traffic congestion when many such gadgets try to interact with each other. The information packets share by the associated gadget on the network in the IoT framework has little payloads and then, due to congestion, packet misfortune results in expensive retransmissions leading to additional delays and enormous overheads. In addition, IoT networks have an associated portability factor because they usually do not handle static nodes. Together, the small packet size, exorbitant re-transmission, and flexibility make the concept of these forced networks completely distinct from the periodic networks that are currently being set up.

Some components have been organized and managed either efficiently or inactively using a different part of the network to regulate congestion. Loss of packets is among the most commonly used congestion indications[15]. When a source sends a packet, it hangs tightly by keeping a clock at its end for a fixed measure of time for a goal affirmation. If no affirmation is obtained within this duration, its packet will be considered lost due to network congestion[16][17]. The congestion control algorithms set up assumes accountability for reducing network congestions whenever a packet loss is differentiated. Instrument that rely on packet losses to differentiate network congestions prove it is of no use in IoT network since: (I) these component expect packet loss occurs due to congestions that is not valid for lower power and loss

978-1-7281-2791-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

network, e.g. packets loss may occur due to link error or weak signal quality [3], and (ii) delaying the congestion reaction until the packet loss occurs.

The study of this paper has the following structure: Section (II) gives information to congestions control& surveyed in different congestion control categories. Section (III) presents analysis of various approaches based on their usage and major concerns. Future scope & conclusion are given in Section (IV).

II CONGESTION CONTROL

Congestion control mechanisms have been generated for several different stages in IoT. Congestion can occur due to high request traffic as well as retransmissions in the network. Retransmission is a common process or an issue for any wireless, low power and lossy network and in that IoT is one of them. Major work is done to control congestion at the protocol level at the application layer and the network layer. This paper analyzes different approaches of the authors to control congestion at the application layer by modifying and updating the constrained application protocol (CoAP) as well as the brief work analysis of the network layer.

Data transmission has increased from the last few years as everyone is well connected. In such situations traffic generated by users is overloaded and it causes packet loss. Offloading is the technique in which the node with low traffic shares the load of a congested node such that the packet will reach within its timeline to reduce the rate of retransmission.

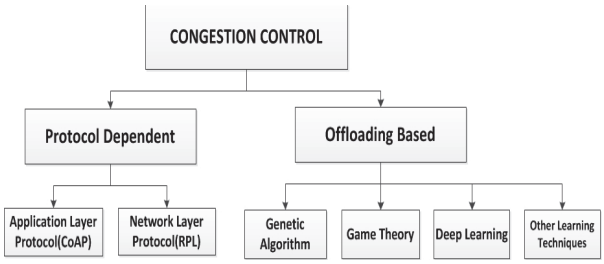


Fig. 1 Categorization of congestion control methods

In protocol based congestion control process value of RTTand RTO are focused to reduce network traffic and avoid congestion whereas, in the process of offloading, the load of the congested nodes is distributed evenly within the network so that traffic can be passed smoothly and reduce the delay in the network.

Congestion control majorly categorized as

- Protocol dependent Congestion control
- Offloading based control congestion

A. PROTOCOL DEPENDENT CONGESTION CONTROL

a) Application Layer Congestion Control Protocol

IoT uses the standard application protocol CoAP for constrained application. This protocol is mainly designed to work over UDP. Various authors show their interest to generate congestions control mechanism for CoAP.

In [19] a thorough study on the congestions control system for CoAP is provided by the author. In this survey majority of work shown is based on the modification of the estimation process of RTO (Retransmission time out). It improves the performance of CoAP/CoCoA but detection of congestion signal from noisy RTT (Round Trip Time) sample is not extracted. There are several studies which compare CoAP/CoCoA[19-21]. CoCoA is considered better than CoAP in some studies whereas some mentioned that CoCoA increases the number of retransmission when the count of the request get increased [24].

In [25] The RTO value is calculated based on the Eifel Re-transmissions timer, research initially suggested for TCP[26]. In the above paper author replace the RFC6298 coefficient α,β, K value with a single coefficient γ value to make it suitable for large sender load. This new variable γ denotes the Ratios of the present RTT and RTO samples. Congestion signal extraction from the RTT sample is missing. In [27] four modification of CoCoA are disused named as (CoCoA-F), CoCoA-strong's (CoCoA-S), and CoCoA-4-Strongs. It works underscores the achievement in presence lossy wireless link of CoCoA and overcomes the side effect of week estimator. CoCoA-F increases competitiveness by reducing the value of the backoff threshold, RTO maximum, and initial values. CoCoA-F also results in a conservative performance like CoCoA. The author subsequently works to differentiate wireless link losses and congestion losses by using 4 state higher granularity estimators. The work increases the performance of CoCoA in the lossy wireless network. CoCoA-4-State-Strong increases the throughput by 30-60% and adapts packet losses. However, the retransmission rate gets increased by around 20%.

In [28] author proposed the adaptive mechanism to handle the congestion in CoAP, this is done by considering packet loss rate and traffic priority. There are major three Components as:- prioritization of congestion, optimized RTO and Return timers. Traffic priorities are allocated according to the type of equipment. Value of RTO backoff is decided based on the traffic priority and packet loss rate. The author doesn't implement the approach which raises the question of its development feasibility. Another approach based on adaptive congestion control mechanism is presented in [29]. This approach is suitable for highly congested network and where a single packet requires multiple retransmissions. This approach precisely selects the value of RTT of the retransmitted packet. The number of successful transaction and throughput is increased but it is not efficient when multiple retransmissions are not required. In [23] author proposed the precise CoCoA based on the limitation of the previous version of CoCoA (CoCoA+): the value of RTO and RTT are too close, weak estimator weights and updates, RTT decrease results in the growth of RTO. The author analyzed the performance of CoCoA in bursty traffic pattern and the result is not good. CoCoA eliminates spurious retransmission and use of weak estimators as well as propose the new method to initialize the RTT measurement parameter. However paper introduces some fixed values which are not practically suited large IoT network scenarios.

In [23] the author introduces the problem of queuing delay in the network termed as buffer bloat. Various congestion

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

control mechanism fails to handle the buffer bloat problem and cause unnecessary retransmission. CoAP and CoCoA are not able to handle this problem and cause wastage of network bandwidth. Author identify the cause of the problem is the inaccuracy of backoff logic of RTO and propose the new backoff logic to overcome the buffer bloat issue.

In [30] also author attempt to overcome the issue of buffer bloat in a heavily congested network. In this Fast- Slow RTO is suggested to determine if the losses of the packets are due to a loss of the wireless link or congestions. There are three features of this: slow RTO and quick RTO computing and self-adaptive re-transmissions timer. Quick RTO is comparable to TCP RTO except for lowest possible RTO; slow RTO function is similar to Karn's algorithm. Fast slow RTO results in shorter Flow Completion time as compare to CoCoA and CoAP as well as in heavily congested traffic it controls RTO. In [31] the author evaluates CoAP congestion control by using Wishful in real-time. Wishful is large scale platform to perform and control run time experiments and also allow creating the unique design of a network. CoAP congestions control mechanism is likened to easy RTT-based algorithms in this document by default. The outcome demonstrates that perhaps the default congestions control mechanism is much more robust and extensible than the straightforward RTT-based algorithms. Vishal Rathod et al [18] have presented a congestion control scheme based on delay gradient for IoT network. The authors aimed to compromise the congestion in the IoT network. To achieve their aim, they have proposed a novel congestion control algorithm which was named as CoCoA++. The authors have attained better network congestion by utilizing delay gradients and a probabilistic backoff. These proposed factors were integrated with CoAP (Constrained Application Protocol). This proposed approach was implemented in the Cooja network simulator. By presenting these approaches, they have achieved a better packet sending rate and delay.

All the above approaches which are discussed in this section are designed for CoAP congestion control mechanism. RTO and RTT begin important factors whose values are responsible to handle and avoid congestion.

b) Network Layer Congestion Control Protocol

Designing the routing protocol in a load aware manner helps in controlling the congestion. Hop to hop communication is considered for congestion control but it increases the overhead as control information is shared from one node to other.

Analysis of 6LoWPAN concerning congestion specifies that packet loss is comparatively high due to buffer overloading instead of channel loss [32]. To control this problem author in [33] proposed a low power and lossy network-based congestion control IPV6 based routing protocol (RPL), which control the packet loss in the congested network due to buffer overflow. The author introduced two different metrics for buffer occupancy and congestion aware objective function which results in the improvement of throughput and packet delivery ratio. In [32] on the other hand, author propose that the congestion control should be measured by using the occupancy of the buffer. This results in the more overhead to transmit information of buffer occupancy. According to [34, 35] during the time of parent selection in routing process mostly packet losses are occurred throughout high traffic scenario due to congestions and load balancing problems. To overcome this problem author in [37] proposed the method termed as simple queue usage basis on the RPL(QU-RPL). In QU-RPL, neighbor node utilization and hop distance from the RPL router is considered to select the parent node. This helps in reduction of High queue loss and a better shipping ratio for packets.

This is useful to regulate transmission energy and route topology jointly and alternately in either the wireless networks to obtain good bandwidth and reliability in the networks [36]. Power controlled RPL is proposed to handle the fluctuation in transmission power to avoid the design of routing topology causes bandwidth loss. It efficiently handles routing topology and transmission power. MarjanFarahani and Akbar GhaffarpourRahbar [46] have presented an energy-efficient load balancing scheme in Dense IoT Networks. The authors aimed to enhance the routing performance of IoT network. To obtain this aim, they have proposed a Double Leveled Unequal Clustering routing protocol. Using this protocol, the load of the traffic has been shared to cluster members. Also, this novel clustering scheme created clusters with unequal-size. Besides, the number of cluster heads was reduced. This proposed approach was simulated in the NS2 network simulator. By presenting this proposed approach, they have increased the throughput as well as they reduced energy consumption.

B. CONGESTION CONTROL BY PERFORMING OFFLOADING

Offloading is a process where a load of traffic node is distributed or shifted to other nodes i.e. pass the load. This process is useful to handle the congestion and reduce delay as well as increase the success rate. Smoothing the traffic in the network extra queues and processing time is included in gateways and other infrastructure which reflects the in-service delay. In either the congested network, requests offloading are proposed to prevent communications failure and useless loss of packets.

In [37] the author proposes the game theory-based computation offloading method to improve benefits for users and reduce the operational cost. The rate of resource allocation at fog and cloud is improved by offloading process. Offloading process optimizes power exploitation which results in energy and delay. In [38] to obtain energy conservations in IoT systems embedded to wireless networks, the author applies cooperative offloading. This system distributes the download process to edge, cloud and mobile nodes. To obtain the best energies efficiency in the IoT network, it utilizes standard various access communications technology. For dense IoT network author in [39] propose offloading method for edge computing exclusively. This approach work in a greedy manner, it consists of 2 tier named as local and mobile edge, among them the incoming task is distributed. It minimizes overhead processing time and energy consumptions. For service-oriented IoT network, author in [40] introduced a joint offloading optimization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

method named as Stabilized green cross haul orchestration. The target of this approach is to enhance energy efficiency and network stability as well as reduce latency. Lyapunoy-theory is used for drift and penalty policies for offloading to determine data processing rate in an energy-efficient manner.

In [41] author proposed the lightweight request framework to improve scalability. It integrates various different architectures and became best suitable for IoT and cloud edge-based integrated architecture. It offers selective independent offloading at the IoT and cloud layer. In [42] author uses both offloading and non-offloading device in the network. The author analyzed modeling and deployment of heterogeneous mobile cloud. Cloud system is analyzed using a remote system whereas others are analyzed by stochastic offloading method. To minimize the outage in IoT communication, it is analyzed the proper placement of cloudlets and offloading process distribution is beneficial. In [43] author tries to improve the IoT device communication using decision offloading with K-means algorithm. to balance the offloading process network get segregated to position edge sever using K-means algorithm. Proposed work reduces the latency and operational cost by using decision-based offloading. In [44] aura perform task offloading for users using IoT based cloud which deliver services in ad-hoc bases. From distributed location aura provide migration, process, and initialization with the help of localized IoT devices. Cost and capacity are managed by this model for load balancing.

The task distribution among edge node based on the constructive syncing method is accessible in [45] toIoT based network. In this approach, tasks are distributed based on the popularity of the process among fog connected edge nodes. Task distribution is based on the one to one game model from cloudlets. Offloading and queuing techniques are used to simultaneously to reduce delay. Hye-Young Kim [52] has presented a load balancing method using loadbot in IoT. The huge amount of demand causes traffic congestion on somespecific paths and the nodes in the IoT network. So, the authors aimed to reduce traffic congestion by solving the problem of IoT demands. To solve this issue, they have proposed an agent, Loadbot that calculated the load of the network and performed structural configurations by examining the number of network loads and user data. Also, for effective load balancing in IoT, they have applied the Deep Belief Network scheme. This proposed approach was implemented in the platform of Matlab. . By presenting this proposed approach, the authors have decreased the network load when the size of network increases. Hye-Young Kim and Jong-Min Kim [47] have presented deep-learning-based load balancing strategy for IoT. Due to the generation of a large number of data over a while in IoT, communications among the users are overloaded. So, the authors aimed to balance the load in IoT. To achieve their aim, they have proposed an agent Loadbot that computed load of the network and have processed the structural configuration using the measured load of network and the number of user data. Also, they have applied the Deep Belief Network scheme for achieving effective load balancing for IoT. Besides, they have presented an agent Balancebot for executing the Q-learning as well as neural prior ensemble-based neural load prediction algorithm. These proposed schemes were implemented in the platform of Matlab. By presenting these approaches, the authors have enhanced the performance of IoT system.

SoulmazGheisari and Ehsan Tahavori [48] have proposed a cognitive method for congestion control in IoT. The authors have aimed to control congestion which occurred because of the large scale of communication and least bandwidth in lossy and low powered networks based IoT. They have attained their aim by including cognition into IoT with the cognitive system based on learning automata. Depend on this system, the authors have a novel cognitive scheme for congestion control with a game of Learning Automata which was abbreviated as CCCLA. Learning Automata was used to control the parameters which can reduce the performance of congestion control. This proposed approach was simulated in the platform of Cooja simulator. By implementing this proposed approach, they have attained better reliability, delay, and throughput. Sungwook Kim and Youngjae Park [49] have proposed data offloading method based on the game for IoT system. The objective of this literature was to overcome the problem of traffic overload that occurred due to a large number of traffic demands. To attain this aim, they have presented a data offloading method based on a novel traffic control strategy. In this approach, they have utilized the Rubinstein bargaining game model and Vickrey-Clarke-Groves (VGC) scheme for reducing the traffic congestion in IoT. Also, they have proposed a novel pricing scheme by assuming the attributes of access point owners and Internet of things modules. This pricing approach has induced the relation between access point owners and Internet of things modules in the process of data offloading. By presenting this proposed approach, the authors have enhanced the quality-of-service.

P. Kuppusamy, R. Kalpana1 and P. V. Venkateswara Rao [50] have proposed data processing and traffic control based on optimization algorithm in IoT. Due to the numerous smart objects and sensors, execution time and human resource were decreased so that the authors have aimed to enhance the processing time of the traffic signal. To achieve this aim, they have proposed a new smart traffic control system which utilized remote cloud server and local traffic smart server. This proposed scheme has tracked the transition of a vehicle that is utilized for capturing high-speed vehicles. For gathering multi-path data, they have presented the Optimized regression algorithm. Also, at four-direction roadway intersection, they have calculated single-point nifty decision with the waiting vehicle density. The result from the simulations showed that the suggested strategy lowered the wait time. Azham Hussain *et al* [51] have presented an adaptive offloading scheme based on Genetic algorithm in IoT devices. The authors aimed to enhance the communication of IoT as it has a huge number of mobile devices and also they have aimed to increase the capacity of traffic handling in IoT. So, for efficient traffic handling, they have proposed adaptive offloading based on a genetic algorithm which was abbreviated as GA-OA. Unessential delays in request process were mitigated by developing the performance of offloading as well as IoT requests success rate was improved. This proposed approach was simulated in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

the platform of Opportunistic Network Environment simulator which was abbreviated as one. Due to the presence of this proposed approach, they have achieved lesser delay, processing time and complexity.

III ANALYSIS OF DIFFERENT APPROACHES

To avoid congestion in constrained IoT networks, various authors have proposed their approaches using different techniques. Congestion control by using the application layer based protocol is important as the proposed approach work properly with its co-existence of standard protocol. Generally in constrained based environment CoAP protocol is preferred, so it is important that all the modifications must be compatible with each other. All approaches must result in high efficient performance under different traffic situations.In routing based congestion control protocol design need to consider energy efficiency while maintaining load. Routing itself consume high amount of energy in forming cluster or selecting the parent node. In [46] author doesn't consider the mobility factor for clustering which is significant for the measurement of the energy loss.

To perform offloading using different learning mechanism is covered. The focus of different approaches is to distribute the traffic evenly among the network such that delay will be reduced. If the count of queries is increased, efficiency can be compromised [53]. It is important to classify and identify the type of packet and identify congestion on that base. Response time must be the base to select and distribute load among neighbors. The main focus of these algorithms is to minimize delay, processing time and complexity. A three proposed solution has been applied for load balancing under heavy and highly dynamic load [52] & it may cause high complexity in IoT network.

IV CONCLUSION AND FUTURE SCOPE

Congestion control is an important issue to be considered while working in IoT networks. There are major 2 ways to control congestion which are, either making the application or network layer or both so efficiently that it can handle congestion in the network as well as control congestion occurred due to reparative re-transmission of the packets. Another method is to offload the congested node and distribute their traffic to neighboring nodes based on their load. To perform offloading, machine learning algorithm has been created to react proactively and control congestion. It is important to identify whether a packet is creating congestion or not so different type of packets should be handled separately so that the delay will be reduced in the processing of the packet and in turn increase the success rate. In the future, it can be planned to propose intelligent learning algorithm to perform offloading as well as control the retransmission rate efficiently. The proposed work should also be adaptive for heterogeneous network with extended flexibility and scalability.

REFERENCES

[1] Y. Sun, Y. Xia, H. Song and R. Bie, Internet of Things Services for Small Towns,in: 2014 International Conference on Identification, Information and Knowledge in the Internet of Things, 2014.

[2] Alzahrani, M. Salha, Sensing for the Internet of Things and its applications, in: 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), pp. 88-92. IEEE, 2017.

[3] Agrawal, Prianka, and Gaurav Chitranshi., Internet of Things for monitoring the environmental parameters, in: 2016 International Conference on Information Technology (InCITe)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds, pp. 48-52. IEEE, 2016.

[4] Ahlgren, Bengt, Markus Hidell, and Edith C-H. Ngai., Internet of things for smart cities: Interoperability and open data, IEEE Internet Computing 20 (6) (2016) 52-56.

[5] Dlodlo, Nomusa, Oscar Gcaba, and Andrew Smith., Internet of things technologies in smart cities, in: 2016 IST-Africa Week Conference, pp. 1-7. IEEE, 2016.

[6] P. Sadhukhan, An IoT based Framework for Smart City Services, in: 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT), 2018.

[7] Stojkoska, BiljanaRisteska, and KireTrivodaliev, Enabling internet of things for smart homes through fog computing, in: 2017 25th Telecommunication Forum (TELFOR), pp. 1-4. IEEE, 2017.

[8] R. Deore, V. Sonawane and P. Satpute, Internet of Thing Based Home Appliances Control,in: 2015 International Conference on Computational Intelligence and Communication Networks (CICN), 2015.

[9] Govindraj, Vignesh, MithileyshSathiyanarayanan, and BabangidaAbubakar, Customary homes to smart homes using Internet of Things (IoT) and mobile application, in: 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), pp. 1059-1063. IEEE, 2017.

[10] Rauscher, Julia, and Bernhard Bauer, Safety and Security Architecture Analyses Framework for the Internet of Things of Medical Devices, in: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1-3. IEEE, 2018.

[11] Laplante, Phillip A., and Nancy Laplante, The internet of things in healthcare: Potential applications and challenges, It Professional 18 (3) (2016) 2-4.

[12] Chiuchisan, Iuliana, Iulian Chiuchisan, and Mihai Dimian, Internet of Things for e-Health: An approach to medical applications, in: 2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM), pp. 1-5. IEEE, 2015.

[13] Gupta, KrishnaKanth, and Sapna Shukla, Internet of Things: Security challenges for next generation networks, in: 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH), pp. 315-318. IEEE, 2016.

[14] Almeida, AF Virgilio, Doneda Danilo, and Monteiro Marilia, Governance challenges for the Internet of Things, IEEE Internet Computing 19 (4) (2015) 56-59.

[15] Halim, NurulHamimiBt, NaimahBtYaakob, and Azmi Bin AwangMd Isa., Congestion control mechanism for Internet-of-Things (IOT) paradigm, in: 2016 3rd International Conference on Electronic Design (ICED), pp. 337-341. IEEE, 2016.

[16] Pan, Yue, Yue Li, and Junxing Zhang, Congestion-aware data acquisition for Internet of Things, in: Proceedings of 2014 International Conference on Cloud Computing and Internet of Things, pp. 131-134. IEEE, 2014.

[17] K. Bhandari, A. Hosen and G. Cho, CoAR: Congestion-Aware Routing Protocol for Low Power and Lossy Networks for IoT Applications, Sensors 18 (11) (2018) 3838.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

[18] V. Rathod, N. Jeppu, S. Sastry, S. Singala and M. Tahiliani, CoCoA++: Delay gradient based congestion control for Internet of Things, Future Generation Computer Systems 100 (2019) 1053-1072.
[19] A. Pramanik, A.K. Luhach, I. Batra, U. Singh,A systematic survey on congestion mechanisms of CoAP based Internet of Things, in: Advanced Informatics for Computing Research, Springer, pp. 306–317, 2017.
[20] A. Betzler, C. Gomez, I. Demirkol, Evaluation of advanced congestion control mechanisms for unreliable CoAP communications, in: Proceedings of the 12th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks, ACM, pp. 63–70, 2015.
[21] E. Ancillotti, R. Bruno, Comparison of CoAP and CoCoA+ congestion control mechanisms for different IoT application scenarios, in: IEEE Symposium onComputers and Communications, ISCC, IEEE, pp. 1186–1192, 2017.
[22] H.M. Hasan, A.I. Ahmed, A comparative analysis for congestion mechanism in COAP and COCOA, Eng. Technol. J. 36 (8 Part (A) Engineering) (2018) 867–877.
[23] S. Bolettieri, C. Vallati, G. Tanganelli, E. Mingozzi, Highlighting some shortcomings of the CoCoA+ congestion control algorithm, in: International Conference on Ad-Hoc Networks and Wireless, Springer, pp. 213–220, 2017.
[24] E. Balandina, Y. Koucheryavy, A. Gurtov, Computing the retransmission timeout in CoAP, in: Internet of Things, Smart Spaces, and Next Generation Networking, Springer, pp. 352–362, 2013.
[25] R. Ludwig, K. Sklower, The Eifel retransmission timer, ACM SIGCOMM Comput. Commun. Rev. 30 (3) (2000) 17–27.
[26] R. Bhalerao, S.S. Subramanian, J. Pasquale, An analysis and improvement of congestion control in the CoAP Internet-of-Things protocol, in: Consumer Communications & Networking Conference (CCNC), 13th IEEE Annual, IEEE, pp. 889–894, 2016.
[27] R. Hassan, A.M. Jubair, K. Azmi, A. Bakar, Adaptive congestion control mechanism in CoAP Application Protocol for Internet of Things (IoT), in: Signal Processing and Communication (ICSC), 2016 International Conference on, IEEE, pp. 121–125, 2016.
[28] J.J. Lee, S.M. Chung, B. Lee, K.T. Kim, H.Y. Youn, Round trip time based adaptive congestion control with CoAP for sensor network, in: International Conference on Distributed Computing in Sensor Systems, DCOSS, IEEE, pp. 113–115, 2016.
[29] I. Järvinen, I. Raitahila, Z. Cao, M. Kojo, Is CoAP congestion safe? in: Proceedings of the Applied Networking Research Workshop, ACM, pp. 43–49, 2018.
[30] C. Vallati, F. Righetti, G. Tanganelli, E. Mingozzi, G. Anastasi, ECOAP: Experimental assessment of congestion control strategies for CoAP using the WiSHFUL platform, in: IEEE International Conference on Smart Computing, SMARTCOMP, IEEE, pp. 423–428, 2018.
[31] H.A. Al-Kashoash, Y. Al-Nidawi, A.H. Kemp, Congestion analysis for low power and lossy networks, in: Wireless Telecommunications Symposium, WTS, IEEE, pp. 1–6, 2016.
[32] H.A. Al-Kashoash, Y. Al-Nidawi, A.H. Kemp, Congestion-aware RPL for 6LoWPAN networks, in: Wireless Telecommunications Symposium, WTS, IEEE, pp. 1–6, 2016.
[33] H.-S. Kim, H. Kim, J. Paek, S. Bahk, Load balancing under heavy traffic in RPL routing protocol for low power and lossy networks, IEEE Trans. Mob.Comput. 16 (4) (2017) 964–979.
[34] H.-S. Kim, J. Ko, D.E. Culler, J. Paek, Challenging the IPv6 routing protocol for low-power and lossy networks (RPL): A survey, IEEE Commun. Surv. Tutor. 19 (4) (2017) 2502–2525.
[35] H.-S. Kim, J. Paek, D.E. Culler, S. Bahk, Do not lose bandwidth: Adaptive transmission power and multihop topology control, in: 13th International Conference on Distributed Computing in Sensor Systems, DCOSS, IEEE, pp. 99–108, 2017.
[36] H. Shah-Mansouri, V. W. S. Wong, Hierarchical fogcloud computing for IoT systems: A computation offloading game. IEEE Internet of Things Journal, 5 (4) (2018) 3246–3257.
[37] H. Guo, J. Liu, H. Qin, Collaborative mobile edge computation offloading for IoT over fiber-wireless networks. IEEE Network, 32(1) (2018) 66–71.
[38] H. Guo, J. Liu, J. Zhang, W. Sun, N. Kato, Mobileedge computation offloading for ultradense IoT networks. IEEE Internet of Things Journal, 5(6) (2018) 4977–4988.
[39] N.-N. Dao, D.-N. Vu, W. Na, , J. Kim & S. Cho, SGCO: Stabilized green crosshaul orchestration for dense IoT offloading services. IEEE Journal on Selected Areas in Communications, 36(11) (2018) 2538–2548.
[40] X. Lyu, H. Tian, L. Jiang, A. Vinel, S. Maharjan, S. Gjessing, et al., Selective offloading in mobile edge computing for the green Internet of Things. IEEE Network, 32(1) (2018) 54–60.
[41] H.-S. Lee, J.-W. Lee, Task offloading in heterogeneous mobile cloud computing: Modeling, analysis, and cloudlet deployment. IEEE Access, 6 (2018) 14908–14925.
[42] C. Zhang, H. Zhao, S. Deng , A density-based offloading strategy for IoT devices in edge computing systems. IEEE Access, 6, 73520–73530.
[43] R. Hasan, M. Hossain, R. Khan, Aura: An incentivedriven ad-hoc IoT cloud framework for proximal mobile computation offloading. Future Generation Computer Systems, 86 2018 821–835.
[44] M. S. Elbamby, M. Bennis, W. Saad, Proactive edge computing in latency-constrained fog networks. In 2017 European conference on networks and communications (EuCNC) (2017).
[45] M. Farahani and A. Ghaffarpour Rahbar, "Double Leveled Unequal Clustering with Considering Energy Efficiency and Load Balancing in Dense IoT Networks", Wireless Personal Communications, vol. 106, no. 3, pp. 1183-1207, 2019.
[46] H.-Y. Kim, A load balancing scheme with Loadbot in IoT networks. The Journal of Supercomputing, 74(3) (2017) 1215–1226.
[47] S. Gheisari and E. Tahavori, CCCLA: A cognitive approach for congestion control in Internet of Things using a game of learning automata, Computer Communications, 147 (2019) 40-49.
[48] Y. Park and S. Kim, Game-based data offloading scheme for IoT system traffic congestion problems, EURASIP Journal on Wireless Communications and Networking, 2015(1) (2015) 1-10.
[49] P. Kuppusamy, R. Kalpana and P. Venkateswara Rao, Optimized traffic control and data processing using IoT, Cluster Computing, 22(1) (2018) 2169-2178.
[50] A. Hussain, S. Manikanthan, T. Padmapriya and M. Nagalingam, Genetic algorithm based adaptive offloading for improving IoT device communication efficiency, Wireless Networks, (2019).
[51] H. Kim, A load balancing scheme with Loadbot in IoT networks, The Journal of Supercomputing, 74(3) (2017) 1215-1226.
[52] S. Taghizadeh, H. Bobarshad and H. Elbiaze, "CLRPL: Context-Aware and Load Balancing RPL for Iot Networks Under Heavy and Highly Dynamic Load", IEEE Access, vol. 6, pp. 23277-23291, 2018.
[53] Mogi, Ryuta, Taichiro Nakayama, and Takuya Asaka. "Load balancing method for IoT sensor system using multi-access edge computing." In 2018 Sixth International Symposium on Computing and Networking Workshops (CANDARW), pp. 75-78. IEEE, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:51:10 UTC from IEEE Xplore. Restrictions apply.

European Journal of Sustainable Development Research
2020, 4(3), em0124
e-ISSN: 2542-4742
https://www.ejosdr.com/

Research Article

MODESTUM
OPEN ACCESS

Biofuel Certification Performance: A Review & Analysis

Sanjay Mohite [1]*, Sagar Maji [2]

[1] Department of Mechanical Engineering, National Institute of Technology, Thanesar, Kurukshetra, Pin Code 136119, Haryana, INDIA
[2] Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi, Pin Code 110042, INDIA
***Corresponding Author:** smohite001@yahoo.com

Citation: Mohite, S., & Maji, S. (2020). Biofuel Certification Performance: A Review & Analysis. *European Journal of Sustainable Development Research, 4*(3), em0124. https://doi.org/10.29333/ejosdr/7864

ARTICLE INFO

Received: 25 Nov. 2019

Accepted: 11 Mar. 2020

ABSTRACT

The biofuel performance certification's scope is reviewed. An operational definition of biofuel performance certification has been developed. The certification of biofuel performance has been recognized as an effective assessment methodology and tool to manage biofuel consumption and improve biofuel performance systematically. It is found that a biofuel performance certification method is required in biofuel industry and biofuel research to authenticate the biodiesel and its blends for use in diesel engine.

Keywords: biofuel, performance, certification, energy audit

GLOBAL CONSUMPTION OF BIOFUEL

There are various advantages with the use of biofuels such as sustainability, greenhouse gas emissions reduction, regional development, social structure and agriculture. A carbon neutral renewable energy is required to mitigate the greenhouse gas effect (Demirbas et al., 2007). Diesel engine causes harmful emissions which affect ecological system adversely. Alternative sources are being explored to solve this problem. Biodiesel is one of the alternative source (Verma et al., 2015).

The 80% share of primary energy is consumed by fossil fuel globally and out of this, 58% is consumed by the transport industry. The International Energy Agency reported that there was a 23% increase in the consumption of diesel while the increase in the consumption of other petroleum products was 7% from 2000 to 2008 (Ayetor et al., 2015). It is reported that there are 100 billion barrels reserves of petroleum globally and are expected to be exhausted in around 40 years (Orhan et al., 2004). It is stated in the International Energy Agency (IEA) report, that 50% more amount of energy will be required than today in 2030. A growth of 1.8% per year would be estimated for the use of global transport energy from 2005 to 2035. It is also reported that the transport industry is responsible for about 60% oil demand in the world. In the transport sector, 97.6% of the fossil fuel energy consumption is oil and the rest is natural gas. Transport sector causes 30% of total U. S. greenhouse gas emissions and it is the second largest source of GHG emissions after electric power sector in the United States. It is reported that production of biodiesel rised from 15 thousand barrels per day to 289 thousand barrels per day from 2000 to 2008. European Union contributes 85% biodiesel production globally (Atabani et al., 2012).

BIOFUEL PERFORMANCE MEASUREMENT

To promote biofuel performance, efficiency of diesel engine's measures such as performance and emission characteristics, energy management monitoring, modelling and optimization have been carried out worldwide. These measures are inadequate in proper evaluation and certification of energy performance in use of biofuel.

Efficiency, performance and emission characteristics, energy measurement and monitoring are an effective measure to provide efficiency data in the support of reliability and applicability of biofuel. These data are useful:

1. To reduce energy consumption
2. To establish environmental performance goals
3. To understand the requirement of diesel engine

Efficiency and energy use should be measured and quantified to control efficiency and energy use in diesel engine, performance and emission characteristics. Energy consumption and energy savings are judged and evaluated with monitoring. Energy modelling and optimization are also the basis to study biofuel performance and have received researchers attention.

Copyright © 2020 by Author/s and Licensed by Modestum Ltd., UK. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ENERGY AUDIT

Energy audit has been performed for a number of buildings, factories, industries with a purpose to improve efficiency and productivity. The energy audit is used to investigate scope to save energy and recommend measures that would make the diesel engine more efficient. Such energy audit would also emphasize key design features and different aspects of operation for overall benefits. Researchers explained energy audit as a process of evaluation of energy use in a building and identify scope to reduce energy consumption (Thumann et al., 2010). A real energy audit called as feasibility study dealt with the thorough examination of two years of utility bills with mathematical modelling of energy use when the system is supposed to operate properly, routine plant operation's review and list of upgrades with their installation costs and savings in a building (Audin, 2002). A good energy audit is helpful to make important decisions on the investment to achieve the best energy efficiency. The development of energy audit is related to the identification and communication of the best energy conservation measures (Anina et al., 2016). An energy audit is one of the step to increase energy efficiency at the facility level. An energy audit has been conducted in a paper mill plant in Guangdong Province in China to identify scope for energy conservation and opportunities for CO_2 emission reduction (Kong et al., 2013). Different people have many meanings for energy audits. Energy audit focuses on process, operation or equipment for some people. Researchers performed an energy audit which was focussed on brick, refractory and insulation giving its effect on efficiency and reliability of boiler and savings of energy fuel (Bases, 2004). In 1988, the computerised transportation audit program was launched by the energy efficiency branch of Alberta's department of energy. This audit was used to investigate the possibility of savings in energy consumption in the transportation sector. An energy audit of a vehicle is a systematic way to assess the uses of energy. The purpose of the audit program is to identify scope to save fuel. The audit is a first step in a process which comprises of setting priorities to develop an energy management program. In energy audit, maintenance factors which reduce friction losses and the factors which affect engine performance are also calculated (Erkut et al., 1992). Energy audit helps to decrease demand and energy consumption in industry (Bellarmine et al., 1996). The ability of emerging technologies to reduce fuel consumption with emphasis on each loss mechanism was investigated in diesel engines with the application of energy audit and technology forecasting. Improvements in fuel consumption were estimated with the help of future technology in heavy duty diesel engines (Thiruvengadam et al., 2014).

The aim of this paper is to carry out a review of biofuel performance certification to evaluate its capability and applicability to improve performance.In the introductory part, global consumption of biofuel, biofuel performance measurement and energy audit have been reviewed and discussed with a view to clarify research gaps.

Definition of Biofuel Certification Performance

The certification of bioenergy sustainability needs international approach and further harmonisation with global monitoring and control. The biofuel certification is a response which concerns with biofuel sustainability. There are various certification schemes which are related to sustainability in other fields. One of the areas in biomass production is covered by some systems such as agriculture, forest and fair trade. Operational experience and effectiveness provide accurate and deep understanding of the structure of certification systems such as design, implementation and constraints. Biofuels should be made sustainable for transport including sustainability of agriculture and forest. All end uses, should be included as a tool in certification schemes to ensure sustainability. Implementation and controlling the quality of certification schemes is important. The proper enforcement and verification mechanisms of certification results in achievenment of sustainability goals. High standards of reliability, transparency including strong implementation, verification requirements and strong sustainability criteria should be ensured in certification schemes. The sustainability criteria for bioenergy has been established by ISO project committee ISO /TC 248. The committee has mentioned the social, economic and environmental issues of the production, supply chain and bioenergy use and determine sustainability criteria. Various biofuel certification scheme are being implemented for biofuel production (Nicolae et al., 2011).

European council directive put a proposal to promote the development of energy savings and emission mitigation. The certification is identified as an effective tool to evaluate and certify energy performance (Cai et al., 2019). A precautionary step is required to assure awareness of biodiesel development and their performance with complete understanding of what could be worked, what could not and why. It also involves analysis of action to be taken to improve outputs and reduce risks during analysing biofuel role (Goetza et al., 2018).

Biofuel industry has great potential for energy saving and its certification. Therefore, energy performance certification of biofuel will perform an important role to achieve energy conservation. This energy performance certification should comprise of a description of key performance and emission characteristics with energy audit parameters and also give options and information to improve performance. Due to various ambiguities, it is not clarified to how to present information about biofuel performance which results in implementation difficulty in view of the requirements.

Scope of Biofuel Certification Performance

The assessment of biofuel sustainability is required to analyse various heterogeneous opinions to consider their multidimensional impacts. Therefore, ability of biofuel to fulfil one specific objective can not lead to any conclusion about overall sustainability of biofuel, its policy efficiency and its positive or negative impacts (Gino et al., 2017). The depletion of fossil fuel resources and global warming challenges various energy sources in the world. The energy demand is rapidly increasing in developing countries. The sustainable biofuels would be considered to replace fossil fuels. Nowadays, focus is to use sustainable biofuels but the definition and assessment of sustainability is found to be highly complex issues. The compulsory targets and financial incentives have been adopted by various countries to promote biofuels and only a few countries were found to work for

Table 1. Type of certification system

Type of certification system	Abbreviation/specific area	Organization/ System
General certification system	CEM	European Committee for Standardization
General certification system containing project approval carbon credits	CDM	Clean Development Mechanism
General certification system	CREM	Consultancy and Research for Environmental Management
General certification system	Ecolabel	Certification of different products and services
General certification system	EMAS	Eco Management and Audit Scheme
General certification system	ISO	International Standard Organization
Biomass energy certification	EUGENE	European Green Electricity Network
Criteria for Sustainability	Biomass tramsite Groep	Workgroup of the Dutch Ministry of Economy; Development of Criteria for sustainable biomass trade
Discussion of Criteria for sustainable biomass trade	Biotrade Workshop	International workshop 2002
Criteria for Sustainability	GRAN	Report, containing Criteria for sustainable biomass trade
Ecological Criteria for Sustainability	Greenpeace	Environmental NGO
Indicator for sustainable development and Agro-ecological indicators	OECD	Organization for Economic Co-operation and Development
Indicator for Sustainable Livelihoods (SL)	UNDP	United Nations Development Program

biofuel sustainability certification scheme within their policy frame work (Gnansounou, 2011). Intergovernmental panel on climate change (IPCC) and International Energy Agency (IEA) are found to reveal the increasing role of bioenergy to mitigate climate change and coherent policy measures are required for this. The policy objectives are difficult to coordinate because of action in one sector may impact unintended negative output in another. Therefore, bioenergy policy should be modelled with the consideration of wide system and overall impacts of policy interventions (Clancy et al., 2018).

While analysing performance certification of biofuel in diesel engines, these methods and measures are found to be considered helpful to improve performance and mitigate emission. Due to the deficiency of studying performance certification in biofuel industries, this paper review and analyse performance certification from the point of view of performance benchmarking which is more beneficial to strengthen and improve performance management. The analysis of scope and data of biofuel performance is helpful to understand application objective and complex rules. It also determines an important base to acquire database for performance certification.

The classification of performance certification by different studies are found to provide details of their certification categories including energy audit to facilitate their implementation in biofuel use. These studies are found to be useful references for studying biofuel performance certification. The proposed performance certification has wide prospects in biofuel application.

During biofuel performance certification, there are two important questions. One is how to define biofuel performance and another is what to measure in biofuel performance. According to previous studies and to perform a detail analysis, the biofuel performance certification should comprise of important performance and emission characteristics to reveal the most basic biofuel performance. In this study, acquiring biofuel consumption of any system is an important basis to design and develop biofuel performance certification. This paper systematically propose performance certification for biofuel use from various aspects including definition of performance certification, classification of performance certification and application of performance certification. According to the definition of the certification, this is a representation of the integrated information consisting of brake specific energy consumption, heat flow analysis, friction power and smoke emission. Therefore, this section attempts to define and classify concepts of these parameters to develop the certification of biofuel use.

Certification System

It is expected that international biomass trade will significantly increase in near future because of the possibility of lower costs of imported biomass, better supply security through diversification and the support by energy and climate policies of various countries. Lower price is one of the reason for international biomass trade. Energy balances and greenhouse gas balances indicates that international bio energy is against energy loss. Certification system provides procedure to develop various types of quality and sustainability standards such as Eco-label, EMAS, ISO, CEM and CREM etc. Various type of organization and system which analyses particular study is shown in table. A comparative table for an in depth analysis of some of the individual certification scheme via its standard to allow a quick and easy comparison of the results is shown in **Table 1**.

Classification on Biofuel Certification Performance

For biofuel performance certification, the initial step is to consider and determine biofuel performance indicators. The indicators can be selected on the basis of their importance in fuel of diesel engine operation such as brake specific energy consumption, heat flow analysis, friction power, smoke emission. Biofuel performance certification is a concept which affects the biofuel performance and comprises of four kinds of information such as brake specific energy consumption, heat flow analysis, friction power, smoke emission. At present, there is no biofuel performance certification which is universally adopted.

There are various parameters which may be considered in preliminary energy audit as shown in **Figure 1**.

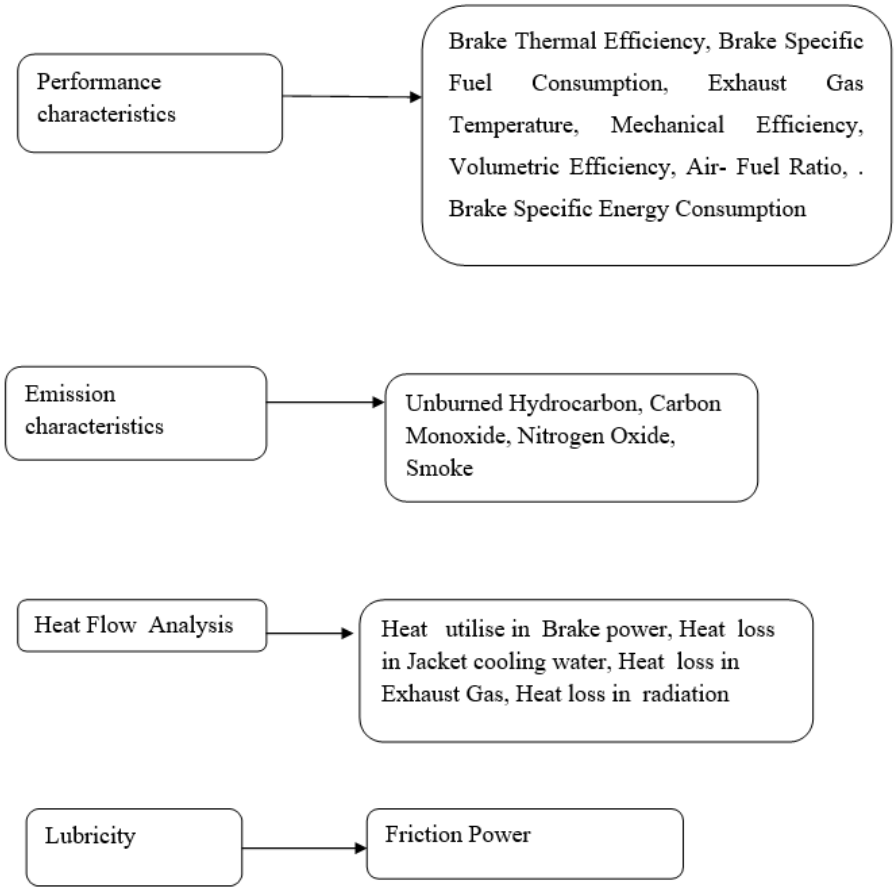


Figure 1. Different parameters for consideration in preliminary energy audit of a diesel engine fuelled with biodiesel

METHODOLOGY

The importance of the selected parameters of the diesel engine fuelled with biodiesel blends has been discussed and their dependency has been found out with parameters. The parameters have been selected on the basis of their influence on other parameters. Brake Thermal Efficiency, Brake Specific Fuel Consumption, Exhaust Gas Temperature, Mechanical Efficiency, Volumetric Efficiency, Air- Fuel Ratio, Unburned Hydrocarbon Emission, Carbon Monoxide Emission, Nitrogen Oxide Emission, Brake Specific Energy Consumption, Heat Flow Analysis, Friction Power, Smoke Emission have been discussed thoroughly. Brake Specific Energy Consumption, Heat Flow Analysis, Friction Power, Smoke Emission have been found to have more influence and effect than other parameters. During primary stage of energy audit, Brake Specific Energy Consumption, Heat Flow Analysis, Friction Power, Smoke Emission have been preferred to compare results (Mohite & Maji, 2020).

DISCUSSION

Brake Specific Energy Consumption

The brake specific fuel consumption shall not be considered as a trustworthy parameter when a blend of two different fuels is prepared because of their different values of calorific values and densities. In case of blending of fuels, brake specific energy consumption will give more correct results. BSEC is one of the key parameters to compute correct energy consumption of biodiesel blends and diesel. It is the product of brake specific fuel consumption and calorific value. Therefore, BSEC is one of the important parameter to evaluate the performance characteristics because it takes into account the calorific value of fuels in blends. BSEC may also be defined as the energy input, which is needed to produce unit brake power. Therefore, BSEC has been chosen as a key parameter for an energy audit of diesel engine fuelled with a biodiesel blend fuel. It also facilitates the proper evaluation of blend fuels as its capability to run a diesel engine Reduction in BSEC is observed with increase in load due to reduction in brake specific fuel consumption at higher loads. Increase in BSEC is observed respectively with the increase in the biodiesel concentration in blend fuel. This may be caused due to dependence of BSEC on BSFC directly as it is a product of calorific value and brake specific fuel consumption (Fattah et al., 2014; Imtenan et al., 2015; Mohite et al., 2016 a). BSFC depends on density and viscosity of fuel and also on brake power. BSEC depends on BSFC and CV as shown in **Figure 2**.

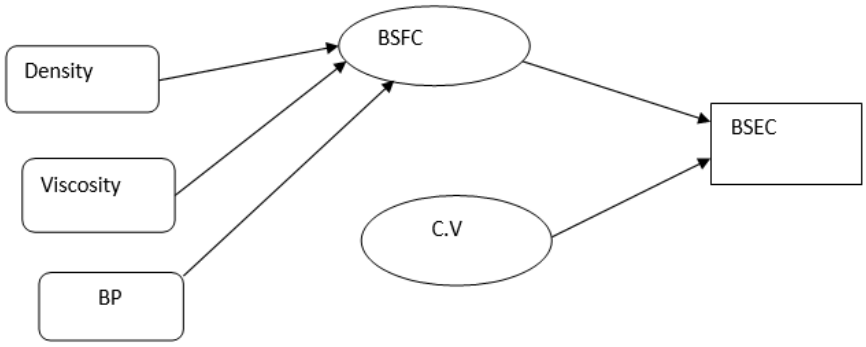


Figure 2. Influence of parameters on BSEC

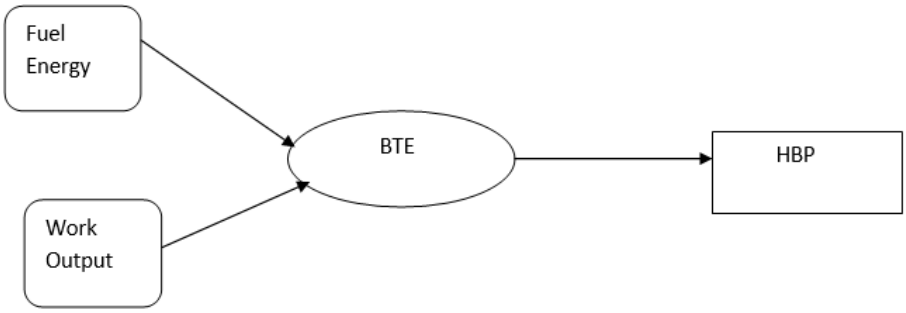


Figure 3. Influence of parameters on HBP

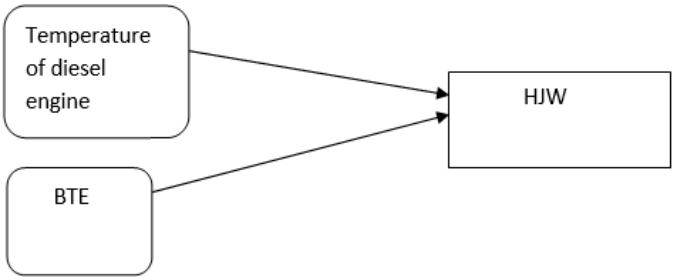


Figure 4. Influence of parameters on HJW

Heat Flow Analysis

Heat flow analysis is one of the key factors to find out the heat energy's utilisation and wastages. Major factors of heat energy consumption are brake power output, coolant medium, exhaust gas and radiation.

Heat Utilise in Brake power

The ratio of work output to the fuel energy supplied is called brake thermal efficiency of an engine. It is also known as fuel conversion efficiency. HBP means brake thermal efficiency as it shows the change of the amount of fuel energy into useful work i.e. brake power output out of total fuel energy. BTE depends on fuel energy and work output as shown in **Figure 3**. BTE of diesel is found to have highest in comparison to all biodiesel blends which is attributed due to lower calorific value of biodiesel (Baiju et al. 2009).

Heat Loss in Jacket Cooling Water

HJW means the amount of part of heat energy which is absorbed by cooling water. This indicates the rise in temperature of combustion chamber, cylinder and the whole diesel engine. It also influence the Brake Thermal efficiency. In water cooled engine, the engine block is surrounded with a jacket of water. A fluid is circulated in jacket of water to engine. About 50% of the friction power is caused between the piston and piston rings and cylinder wall and is transferred to the cooling media as heat energy. The rest of the friction power is caused because of bearings, valve mechanism and to drive auxiliary devices and is changed as heat energy to the oil or surrounding. Heat carried away by the cooling media consist of heat transmitted to the combustion chamber walls from incylinder gases, heat transmitted to the exhaust valve during the exhaust process and a significant fraction of the friction work. At part load, a larger part of the fuel's calorific value is absorbed by the cooling medium. Heat transfer rate in the cooling medium is 2 to 3 times of the brake power at low speeds and loads (Heywood, 2012). HJW depends on temperature of diesel engine and BTE as shown in **Figure 4**.

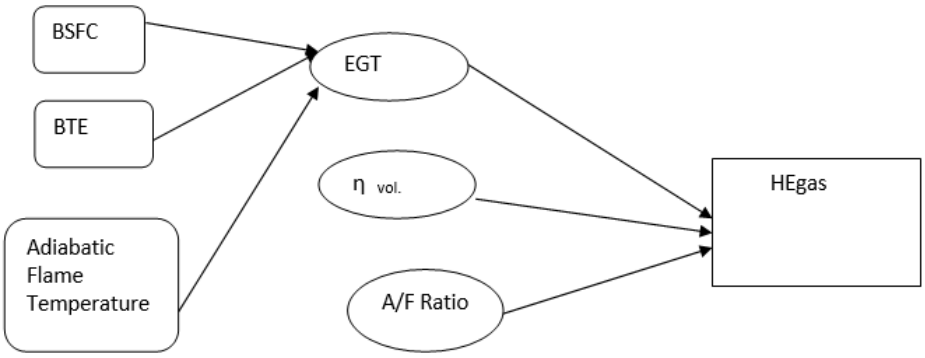


Figure 5. Influence of parameters on HEgas



Figure 6. Influence of parameters on HRAD

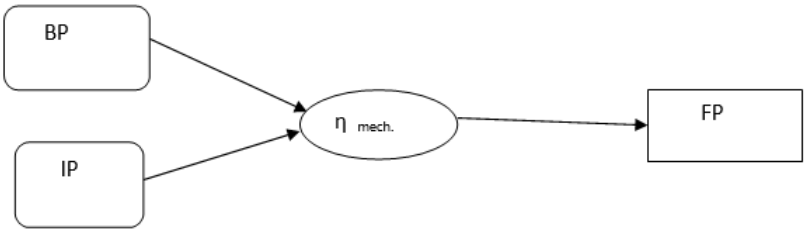


Figure 7. Influence of parameters on FP

Heat Loss in Exhaust Gas

The enthalpy of the exhaust gases may be divided into the following parameter: a 60% sensible enthalpy, 7% exhaust kinetic energy, 20% incomplete combustion and 12% heat transfer to the exhaust system. It is also investigated that some part of heat transfer to the exhaust system is radiated to the surrounding and the rest of it ends up in the cooling medium (Heywood, 2012). This is loss of heat to exhaust gas. Heat to exhaust gas is directly related to exhaust gas temperature. EGT increases with load for all fuels due to the requirement of more power and fuel. Flame temperature of biodiesel is higher than that of diesel at a particular load in spite of the lower calorific value of biodiesel, which causes an increase in EGT. EGT is found to be higher due to higher flame temperature of biodiesel as compared to that of diesel (Mohite et al., 2016b). HEgas depends on EGT, volumetric efficiency and air fuel ratio. EGT depends on BSFC, BTE and adiabatic flame temperature as shown in **Figure 5**.

Heat Loss in Radiation

This is the heat energy which is found to be lost in radiation and other unaccounted losses. Heat transfer due to radiation was found to be varied in the range of 0.5% to 10% of the total chemical energy released during combustion (Benajes et al., 2015). Heat transfer to radiation may also ranges from 20 to 35% of the total heat transfer in diesel engines (Heywood, 2012). HRAD depends on BTE as illustrated in **Figure 6**.

Friction Power

The deduction between indicated power and brake power is called as friction power. This power is found to be in the friction between moving parts of the engine, draw fresh air into the intake system and to the engine accessories driving. Mechanical efficiency is also found to be defined as the ratio of brake power to the sum of brake power and friction power. Lubricity decreases friction power which causes an increase in mechanical efficiency provided brake power is same. FP depends on mechanical efficiency and mechanical efficiency depends on BP and IP as illustrated in **Figure 7**.

Dhar and Agarwal (2014) reported minor wear of valves, pistons, piston rings, liners and small end bearing of connecting rod while using diesel blended with 20% Karanja biodiesel. High lubricity is the main advantage of biodiesel, which reduces friction losses and improve brake power. Good lubrication reduces friction in engine components, thereby reducing energy consumption and wear (Tung & McMillan, 2004).

Smoke Emission

Flame is not uniform in a diesel engine at the time of combustion. During self-ignition, flame covers all areas of combustion chamber which have air-fuel ratio in combustible range. Equivalence ratio ranges from 0.8 to 1.5 which assist in combustion. When there is rich air-fuel mixture in the combustion zone, there is not sufficient oxygen to form carbon dioxide, but carbon monoxide

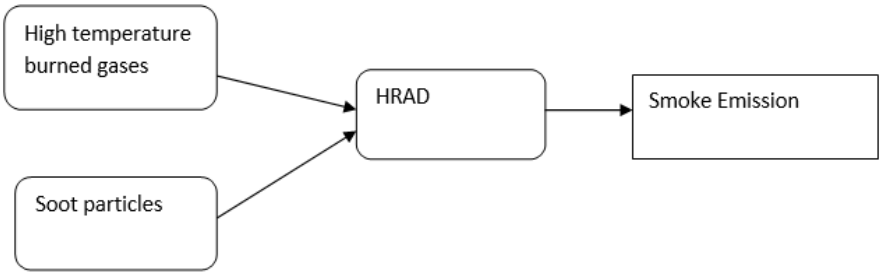


Figure 8. Influence of parameters on Smoke

and solid carbon would be formed. Black smoke comprises of these solid carbon particles. Diesel engines operate with an overall lean air-fuel ratio. Density of particulate emission is found to be maximum when the engine is under higher load. This is due to higher quantity of fuel injected to get more power resulting in rich mixture and poor fuel economy. Soot particles comprises of groups of solid carbon spheres. The diameter of these solid carbon spheres ranges from 10 nm to 80 nm (1nm=10^{-9}m) (Pulkrabek, 2003). Radiation by soot was found to be a significant source of losses in efficiency of engines. The soot was found to be luminous during combustion and it was found to be an important source of radiation. Carbon dioxide and water vapour molecules were also found to emit radiation, but its quantity was found to be lower than that of soot particles (Benajes et al., 2015). The high temperature burned gases and the soot particles are the two sources of radiation in the diesel engine. Fuel burns in a turbulent diffusion flame when fuel and air mix together in diesel engine. The flame is found to be luminous and soot particles which mostly consist of carbon are formed in the combustion process. In a diesel engine, the radiation of soot particles is nearly five times the radiation of the gaseous combustion products (Heywood, 2012). Thus, it is investigated that smoke emission is more harmful to the atmosphere and it also increases heat loss to radiation. Therefore, this characteristic must be chosen as a key parameter in the Energy Audit as compared to other emission characteristics such as CO, UBHC and NO_x. Thus, increase in smoke emission results in increase in heat loss to radiation. Thus, smoke depends on heat radiation and heat radiation depends on high temperature burned gases and soot particles as illustrated in **Figure 8**.

CONCLUSIONS

It is concluded that the following performance and emission characteristics of the diesel engine have been chosen as important parameters to evaluate diesel engine fuelled with diesel and biodiesel blends by energy audit at the primary level (Mohite & Maji, 2020).

1. Brake Specific Energy Consumption
2. Heat Flow Analysis
    a) Heat utilise in Brake power (HBP)
    b) Heat loss in Jacket cooling water (HJW)
    c) Heat loss in Exhaust Gas (HEgas)
    d) Heat loss in Radiation (HRAD)
3. Friction Power
4. Smoke Emission

This may be chosen as preliminary energy audit method to evaluate diesel engine fuelled with biodiesel blends to save time. If the blend fuel is found to be feasible in diesel engine, then it would be used in diesel engine else it would to rejected altogether without the need of evaluation of any other parameters. This may be recognised as a method of certification for biofuel performance in diesel engine.

ABBREVIATIONS

BP	: Brake Power
BSFC	: Brake Specific Fuel Consumption
C.V.	: Calorific Value
BSEC	: Brake Specific Energy Consumption
BTE	: Brake Thermal Efficiency
EGT	: Exhaust Gas Temperature
$\eta_{vol.}$	: Volumetric Efficiency
A/F Ratio	: Air Fuel Ratio
IP	: Indicated Power
$\eta_{mech.}$	: Mechanical Efficiency

REFERENCES

Anina, J. M., & Rottmayer, S. P. (2016). Virtual audits: the promise and the reality. *Energy Engineering*, *113*(6), 34-52. https://doi.org/10.1080/01998595.2016.11772067

Atabani, A. E., Silitonga, A. S., Badruddin, I. A., Mahlia, T. M. I., Masjuki, H. H., & Mekhilef, S. (2012). A comprehensive review on biodiesel as an alternative energy resource and its characteristics. *Renewable and Sustainable Energy Reviews*, *16*, 2070-2093. https://doi.org/10.1016/j.rser.2012.01.003

Audin, L. (2002). How long since your last energy audit. *Strategic Planning for energy and Environment*, 22(1), 72-75. https://doi.org/10.1080/10485230209509606

Ayetor, G. K., Sunnu, A., & Parbey, J. (2015). Performance evaluation of biodiesel -biodiesel blends in a dedicated CIDI engine. *International Journal of Renewable Energy Research, 5*(1), 168-176.

Baiju B., Naik M. K., & Das L. M. (2009). A comparative evaluation of compression ignition engine characteristics using methyl & ethyl esters of karanja oil. *Renewable Energy*, *34*(6), 1616-1621. https://doi.org/10.1016/j.renene.2008.11.020

Bases, G. J. (2004). An energy audit that saved real energy and money. *Proceedings of NAWTEC 12, 12th North American Waste to Energy Conference*, *May 17-19, 20*04, Savannah, Georgia, USA, NAWTEC, 12-2221. https://doi.org/10.1115/NAWTEC12-2221

Bellarmine, G. T., & Arokiaswamy, N. S. S. (1996). Energy management techniques to meet power shortage problems in India. *Energy Conversion Management, 37*(3), 319-328. https://doi.org/10.1016/0196-8904(95)00181-6

Benajes J, Martin J, Garcia A, Villalta D, & Warey A. (2015). Incylinder soot radiation heat transfer in direct injection diesel engine. *Energy Conversion and Management*, 106, 414-427. https://doi.org/10.1016/j.enconman.2015.09.059

Cai, W., Liu, C., Lai, K. H., Li, L., Cunha, J., & Huf, L. (2019). Energy performance certification in mechanical manufacturing industry: A review and analysis. *Energy Conversion and Management, 186*, 415-432. https://doi.org/10.1016/j.enconman.2019.02.041

Clancy, J. M., Curtisd, J., & O'Gallachoira, B. (2018). Modelling national policy making to promote bioenergy in heat, transport and electricity to 2030 - Interactions, impacts and conflicts. *Energy Policy*, 123, 579-593. https://doi.org/10.1016/j.enpol.2018.08.012

Demirbas, A. H., & Demirbas, I. (2007). Importance of rural bioenergy for developing countries. *Energy Conversion and Management*, *48*, 2386-2398. https://doi.org/10.1016/j.enconman.2007.03.005

Dhar, A., & Agarwal, A. K. (2014). Effect of Karanja biodiesel blend on engine wear in a diesel engine. *Fuel*, *134*, 81-89. https://doi.org/10.1016/j.fuel.2014.05.039

Erkut, E., & Maclean, D. (1992). Alberta's energy efficiency branch conducts transportation audits. *Interfaces*, *22*(3), 15-21. https://doi.org/10.1287/inte.22.3.15

Fattah, I. M. R., Masjuki, H. H., Kalam, M. A., Wakil, M. A., Ashraful, A. M., & Shahir, S. A. (2014). Experimental investigation of performance and regulated emissions of a diesel engine with Calophylum inophyllum biodiesel blends accompanied by oxidation inhibitors. *Energy Conversion and Management, 83*, 232-240. https://doi.org/10.1016/j.enconman.2014.03.069

Gino, B., Delrued, F., Legranda, J., Pruvosta, J., & Valléea, T. (2017). The challenge of measuring biofuel sustainability: A stakeholder-driven approach applied to the French case. *Renewable and Sustainable Energy Reviews*, *69*, 933-947. https://doi.org/10.1016/j.rser.2016.11.022

Gnansounou, E. (2011). Assessing the sustainability of biofuels: A logic-based model. *Energy*, *36*, 2089-2096. https://doi.org/10.1016/j.energy.2010.04.027

Goetza, A., Searchinger, T., Beringer, T., German, L., McKay, B., Oliveira, G. L. T., & Hunsberger, C. (2018). Reply to commentary on the special issue Scaling up biofuels? A critical look at expectations, performance and governance. *Energy Policy*, *118*, 658-665. https://doi.org/10.1016/j.enpol.2018.03.046

Heywood, J. B. (2012). Internal Combustion Engine Fundamentals, Tata McGraw Hill Edition.

Imtenan, S., Masjuki, H. H., Varman, M., Fattah, I. M. R., Sajjad, H., & Arbab, M. I. (2015). Effect of n-butanol and diethyl ether as oxygenated additives on combustion emission performance characteristics of a multiple cylinder diesel engine fuelled with diesel - jatropha biodiesel blend. *Energy Conversion and Management, 94*, 84-94. https://doi.org/10.1016/j.enconman.2015.01.047

Kong, L., Price, L., Hasanbeigi, A., Liu, H., & Li, J. (2013). Potential for reducing paper mill energy use and carbon dioxide emissions through plant -wide energy audits: a case study in China. *Applied Energy*, *102*, 1334-1342. https://doi.org/10.1016/j.apenergy.2012.07.013

Lewandowski, I., & Faaij, A. P. C. (2005). An overview on approaches to assess sustainability made from Fair (Bio) trade as preparation for the development of a certification systems for large scake sustainable import of (energy from) biomass. Report NW-SE-2004, Utrecht: University Utrecht, Copernicus Institute, Development of Sciece, Technology and Society; 2003.

Lewandowski, I., & Faaij, A. P. C. (2006). Steps towards the developmentmof a certification system for sustainable bio-trade. *Biomass and Bioenergy*, *30*, 84-104. https://doi.org/10.1016/j.biombioe.2005.11.003

Mohite, S., & Maji, S. (2020). Importance of Energy Audit in Diesel Engine Fuelled with Biodiesel Blends: Review and Analysis. *European Journal of Sustainable Development Research, 4*(2), em0118. https://doi.org/10.29333/ejosdr/7596

Mohite, S., Kumar, S., & Maji, S. (2016a). Performance Characteristics of mix oil biodiesel blends with smoke emissions. *International Journal of Renewable Energy Development*, *5*(2), 163-170. https://doi.org/10.14710/ijred.5.2.163-170

Mohite, S., Kumar, S., & Maji, S. (2016b). Experimental studies on use of karanja biodiesel as blend in a compression ignition engine. *International Journal of Renewable Energy Research*, *6*(2), 355-360.

Nicolae, S., & Jean-Francois, D. (2011). Recent developments of biofuels/bioenergy sustainability certification: A global overview. *Energy Policy*, *39*, 1630-1646. https://doi.org/10.1016/j.enpol.2010.12.039

Orhan, A. S., Dulger, Z., Kahraman, N., & Veziroglu, T. N. (2004). Internal combustion engines fuelled by natural gas-hydrogen mixture. *International Journal of Hydrogen Energy*, *29*, 1527-1539. https://doi.org/10.1016/j.ijhydene.2004.01.018

Pulkrabek, W. W. (2003). Engineering Fundemantals of the Internal Combustion Engine. University of Wisconsin-Patteville, Prentice Hall, Upper Saddle River, New Jersey, 07458.

Thiruvengadam, A., Pradhan, S., Thiruvengadam, P., Besch, M., & Carder, D. (2014). Heavy duty vehicle diesel engine efficiency evaluation and energy audit. *Center for Alternative Fuels, Engines, and Emissions, West Virgina University, Morgantown, WV, Final report on Oct, 2014 published by Oscar Delgado, the International Council on Clean Transportation, Washignton DC.*

Thumann, A., & Younger, W. (2010). Handbook of energy audit, 7th Edition, Lilburn, GA: the Fairmont Press, Inc.

Tung, S. C., & McMillan, M. L. (2004). Automotive tribology overview of current advances and challenges for the future. *Tribology International*, *37*, 517-536. https://doi.org/10.1016/j.triboint.2004.01.013

Verma, P., & Sharma, M. P. (2015). Performance and emission characteristics of biodiesel fuelled diesel engines. *International Journal of Renewable Energy Research*, *5*(1), 245-250.

Natl. Acad. Sci. Lett.
https://doi.org/10.1007/s40009-020-00971-7

SHORT COMMUNICATION

Characterization of Line-Cut Signed Graphs

Rashmi Jain[1] · **Mukti Acharya**[3] · **Sangita Kansal**[2]

Received: 3 November 2017 / Revised: 1 March 2020 / Accepted: 12 March 2020
© The National Academy of Sciences, India 2020

Abstract A *signed graph* $S = (S^u, \sigma)$ consists of an underlying graph S^u and a function $\sigma : E(S^u) \longrightarrow \{+, -\}$. For a graph G, its lict graph written as $L_c(G)$ is the intersection graph $\Omega(E(G) \cup C(G))$. For a signed graph S, its *line-cut signed graph* (in short called lict signed graph) is denoted by $L_c(S)$ has underlying graph $L_c(S^u)$ and two vertices u and v joint by negative edge if u and v both are negative adjacent edges of S or u is a negative edge incident to a cut-vertex v of negative degree odd in S and by positive edge otherwise, here $C(S)$ is the set of cut-vertices of S. In this paper, we establish structural characterization of lict signed graphs $L_c(S)$.

Keywords Signed graph · Lict signed graph · Line signed graph

Mathematics Subject Classification 05C22 · 05C75 · 05C76

✉ Rashmi Jain
rashmi2011f@gmail.com

Mukti Acharya
mukti1948@gmail.com

Sangita Kansal
sangita_kansal15@dce.ac.in

[1] Department of Mathematics, Government College Patharia, Patharia, Dist. Damoh, M.P. 470666, India

[2] Department of Applied Mathematics, Delhi Technological University, Shahbad Daulatpur, Main Bawana Road, Delhi 110 042, India

[3] Department of Mathematics, CHRIST (Deemed to be University), Bengaluru, Karnataka 560029, India

For general notations and concepts in graph theory, we refer to [9]. Unless specified otherwise, all graphs considered here are simple, connected, finite and undirected.

A *signed graph* $S = (S^u, \sigma)$ consists an underlying graph S^u and a function $\sigma : E(S^u) \longrightarrow \{+, -\}$ that is said to be the *signature* of S. A signed graph, in short, is said to be a *sigraph*. Any edge of positive (negative) sign is called a *positive* (*negative*) *edge* and the sets $E^+(S) = \{e \in E(S^u) : \sigma(e) = +\}$ and $E^-(S) = \{e \in E(S^u) : \sigma(e) = -\}$ are the sets of positive and negative edges of S, respectively. Thus, the edge set of S is $E(S) = E^+(S) \cup E^-(S)$. In a signed graph, we denote positive edges by solid lines and negative edges by dashed lines, as shown in Fig. 1. The total degree of v is $d(v) = d^+(v) + d^-(v)$, where $d^+(v)(d^-(v))$ is called positive (negative) degree of v, that is the number of positive (negative) edges incident to vertex v and a vertex of negative degree even (odd) is called a positive (negative) vertex.

Two signed graphs S_1 and S_2 are called *isomorphic* signed graphs, written as $S_1 \cong S_2$, if there underlying graphs are isomorphic and the signs of corresponding edges are same. The *line graph* [9] of a graph G is denoted by $L(G)$ whose vertex set is $E(G)$, in which two vertices are adjacent if the corresponding edges in G are adjacent. Several derived signed graphs like line signed graph [1], product line signed graphs [3], dot line signed graphs [4], 2-path signed graphs [12], 2-path product signed graphs [13], square signed graphs [12, 14, 15], common edge signed graphs [2, 11], signed line digraph [5], and lict signed graphs [7] are studied in the literature. Here, we focus on another beautiful derived structure called line-cut signed graphs. Now, we state the results which are needed for our work.

Published online: 02 April 2020

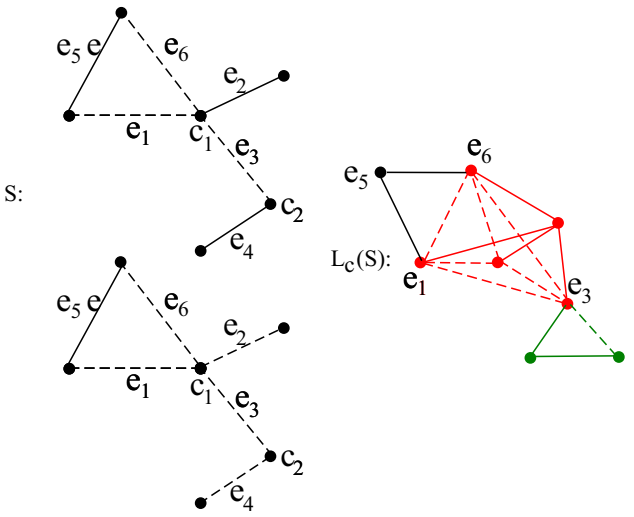


Fig. 1 Two signed graphs and their common lict signed graph

Theorem 1 [9] *A graph is a line graph if and only if its edges can be partitioned into complete subgraphs in such a manner that no vertex lies in more than two of these subgraphs.*

Definition 1 [8] The **line signed graph** $L(S)$ of a signed graph S has underlying graph $L(S^u)$ and edge ee' is negative if e and e' are negative edges of S and positive otherwise.

Definition 2 [10] For a graph G, its **lict graph** is denoted by $L_c(G)$. It has vertex set $E(G) \cup C(G)$ in which two vertices are adjacent if they are two adjacent edges of G or one vertex is an edge e_i of G and the other vertex is a cut-vertex c_j of G such that e_i is incident to c_j, i.e., $L_c(G)$ is the intersection graph $\Omega(E(G) \cup C(G))$. Clearly, $L_c(G) \cong L(G)$ if and only if $C(G) = \phi$.

A graph G is said to be a lict graph if and only $G \cong L_c(H)$ for some graph H.

Remark 1 By the definition of lict graph, two adjacent vertices of $L_c(G)$ cannot be cut-vertices of G .

A *clique* of a graph is its maximal complete subgraph. The following theorem is the characterization of lict graphs:

Theorem 2 [6] *The following statements are equivalent:*

(1) *Graph G is a lict graph.*
(2) *The edges of G can be partitioned into cliques in such a manner that no vertex lies in more than two of these cliques and for each clique G',*
   (i) *if each vertex of G' lies in two cliques of the partition, then $G - E(G')$ is connected; and*
   (ii) *if atleast one vertex v of G' does not lie in another clique of the partition, then $G - E(G') - v$ is disconnected.*

Definition 3 [7] For a signed graph $S = (S^u, \sigma)$, its *lict signed graph* is $L_c(S) = (L_c(S^u), \sigma')$, where for every edge uv of $L_c(S^u)$

$$\sigma'(uv) = \begin{cases} - & \text{if } u, v \in E^-(S) \text{ or } u \in E^-(S),\ v \in C^-(S); \\ + & \text{otherwise.} \end{cases}$$

Here, $C^-(S)$ is the set of cut-vertices of S having negative degree odd.

In Fig. 1, two signed graphs and there lict signed graph are shown. A signed graph S is called a lict signed graph if and only if $S \cong L_c(T)$ for some signed graph T. This signed graph T is called the lict root of S.

Kulli and Muddebihal [10] introduced the concept of line-cut graph in 2006. Acharya et al. [6] characterized the line-cut graphs in 2014 that was long waited. They also introduced line-cut signed graphs in 2015 that is extension of line signed graphs introduced by Behzad and Chartrand [8] in 1969. In this paper, we establish necessary and sufficient conditions for a signed graph to being a line-cut signed graph.

Now, the following is the main result of this paper:

Theorem 3 *S is a lict signed graph if and only if the following conditions are satisfied in S:*

(i) *S^u is a lict graph;*
(ii) *there is no path $P_4 = (u, v, w, x)$ or a triangle (u, v, w) in which only one edge vw is positive and;*
(iii) *for any clique S_i, if some vertices of S_i do not lie in another cliques, then for a vertex (say c), either $d^-(c) = 0$ or odd. If $d^-(c)$ is 0 (odd), then an even number of vertices adjacent to c (end points of all negative edges incident on c) must be adjacent to each other by negative edges.*

Proof *Necessity*:

Suppose S is a lict signed graph then for some signed graph T, $S \cong L_c(T)$. Hence, $S^u \cong L_c(T^u)$, i.e., S^u is a lict graph. Thus, **(i)** follows.

Suppose there exists a path $P_4 = (u, v, w, x)$ in which only one edge vw is positive. Since two adjacent vertices of S are not cut-vertices of T, we find following possibilities for T:

Case **I:** If u, v, w and x are edges of T, then $uv, wx \in E^-(S)$ implies that u, v, w and x are negative edges of T. Hence, vw must be negative edge of S.

Case **II:** If v, w, x are edges of T and u is a cut-vertex of T, then $uv, wx \in E^-(S)$ implies that u is negative cut-vertex of T and v, w, x are negative edges of T. Hence, vw must be negative edge of S.

Case **III:** If u, w are cut-vertices of T and v, x are edges of T, then $uv, wx \in E^-(S)$ implies that u, w are negative cut-vertices of T and v, x are negative edges of T. Hence, vw must be negative edge of S.

Case **IV:** If u, x are cut-vertices of T and v, w are edges of T then $uv, wx \in E^-(S)$ implies that u, x are negative cut-vertices of T and v, w are negative edges of T. Hence, vw must be negative edge of S.

Similarly, if there is a triangle (u, v, w) in S, in which only edge vw is positive, then following are possible cases:

Case **I:** If u, v and w are edges of T, then $uv, wu \in E^-(S)$ implies that u, v and w are negative edges of T. Hence, $vw \in E^-(S)$.

Case **II:** If u is a cut-vertex of T and v, w are edges of T, then $uv, wu \in E^-(S)$ implies that u is negative cut-vertex of T and v, w are negative edges of T. Hence, $vw \in E^-(S)$.

In the above cases, we find that vw is a negative edge of S that contradicts our assumption. Thus, **(ii)** follows.

Next, suppose for any clique S_i of S, some vertices of S_i do not lie in another cliques, then by definition of lict signed graph, one vertex (say c) in such vertices corresponds to a cut-vertex and other vertices correspond to pendant edges of signed graph T.

Now, there are only two possible cases for a cut-vertex c in signed graph T:

Case **I:** If $d^-(c)$ **is even**, then by definition of lict signed graph, $d^-(c) = 0$ in signed graph S and an even number of vertices adjacent to c in S must join to each other by negative edges.

Case **II:** If $d^-(c)$ **is odd,** then $d^-(c)$ must be odd in signed graph S and other end vertices of all negative edges incident to c in S must join to each other by negative edges as they correspond to adjacent negative edges in signed graph T.

Hence, **(iii)** follows.

Thus, the necessity is proved.

Sufficiency:

Let all conditions of the theorem be satisfied for a signed graph S. Now, we give the construction of lict root of S. Since S^u is a lict graph, by Theorem 2, the edges of S^u can be partitioned into cliques in such a manner that no vertex lies in more than two of these cliques. Let $\mathcal{P}(S) = \{S_1, S_2, ..., S_n\}$ be such a partition of $E(S)$. We draw a graph T^u having $V(T^u) = \mathcal{P}(S) \cup U$, where the set U contains those vertices of S which belong only to one of the cliques S_i leaving one such vertex for each S_i, and two of these vertices are adjacent whenever they have a nonempty intersection. Next, we assign the labels to the edges and cut-vertices of T^u and then give labels to vertices of S and construct signed graph $T = (T^u, \sigma)$ such that

$$\sigma(e_i) = \begin{cases} +, & \text{if } d^-(e_i) = 0 \text{ in } S; \\ -, & \text{otherwise.} \end{cases}$$

For this signed graph T, we have $S \cong L_c(T)$. □

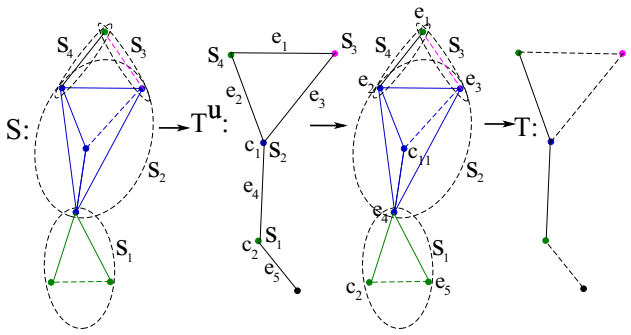


Fig. 2 Construction of lict root T of lict signed graph S

Figure 2 illustrates the construction of lict root of S satisfying sufficiency conditions of Theorem 3 □.

Proposition 4 [1] *S is a line signed graph if and only if S^u is a line graph and there is no path $P_4 = (u, v, w, x)$ or a triangle (u, v, w) in which only edge vw is positive.*

The characterization of product and dot line-cut signed graphs is reported in our subsequent paper.

References

1. Acharya M, Sinha D (2005) Characterizations of line sigraphs. Nat Acad Sci Lett 28(1–2):31–34
2. Acharya M, Sinha D (2006) Common-edge sigraphs. AKCE Int J Graphs Comb 3(2):115–130
3. Acharya M (2009) ×-Line signed graphs. J Comb Math Comb Comput 69:103–111
4. Acharya BD (2010) Signed intersection graphs. J Discrete Math Sci Cryptogr 13(6):553–569
5. Acharya M, Sinha D (2013) Characterization of signed line digraphs. Discrete Appl Math 161(9):1170–1172
6. Acharya M, Jain R, Kansal S (2014) Characterization of line-cut graphs. Graph Theory Notes N Y 66:43–46
7. Acharya M, Jain R, Kansal S (2015) Results on lict signed graphs $L_c(S)$. J Discrete Math Sci Cryptogr 18(6):727–742
8. Behzad M, Chartrand GT (1969) Line coloring of signed graphs. Elem Math 24:49–52

9. Harary F (1969) Graph theory. Addison-Wesley Publ. Comp, Reading
10. Kulli VR, Muddebihal MH (2006) The lict graph and litact graph of a graph. J Anal Comput 2(1):33–43
11. Sinha D, Upadhayaya S, Kataria P (2013) Characterization of common edge signed graphs. Discrete Appl Math 161(9):1275–1285
12. Sinha D, Sharma D (2016) On square and 2-path signed graphs. J Interconnect Netw. https://doi.org/10.1142/S0219265915500115
13. Sinha D, Sharma D (2017) Characterization of 2-path product signed graphs with its properties. Comput Intell Neurosci 2017:1–8. https://doi.org/10.1155/2017/1235715
14. Sinha D, Sharma D (2018) On the properties of square signed graph. Natl Acad Sci Lett. https://doi.org/10.1007/s40009-018-0643-0
15. Sinha D, Sharma D, Square signed graph. Natl Acad Sci Lett. https://doi.org/10.1007/s40009-018-0781-4

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW)

Classification of Cuisines from Sequentially Structured Recipes

1st Tript Sharma
Department of Mechanical Engineering
Delhi Technological University
Delhi, India
triptsharma22@gmail.com

1st Utkarsh Upadhyay
Department of Electronics Engineering
Jamia Millia Islamia University
Delhi, India
utkarshdhy@gmail.com

Ganesh Bagler
Department of Computational Biology
Indraprastha Institute of Information Technology Delhi (IIIT-Delhi)
bagler@iiitd.ac.in

***Abstract*—Cultures across the world are distinguished by the idiosyncratic patterns in their cuisines. These cuisines are characterized in terms of their substructures such as ingredients, cooking processes and utensils. A complex fusion of these substructures intrinsic to a region defines the identity of a cuisine. Accurate classification of cuisines based on their culinary features is an outstanding problem and has hitherto been attempted to solve by accounting for ingredients of a recipe as features. Previous studies have attempted cuisine classification by using unstructured recipes without accounting for details of cooking techniques. In reality, the cooking processes/techniques and their order are highly significant for the recipe's structure and hence for its classification. In this article, we have implemented a range of classification techniques by accounting for this information on the RecipeDB dataset containing sequential data on recipes. The state-of-the-art RoBERTa model presented the highest accuracy of 73.30% among a range of classification models from Logistic Regression and Naive Bayes to LSTMs and Transformers.**

***Index Terms*—Recurrent Neural Networks, Transformers, Classification, Sequential Recipes**

I. Introduction

Cuisines represent the culinary imprint of cultures. The structure of cuisines is shaped by composition of their recipes. Increasing availability of data of cuisines has led to data-driven explorations of cuisines such as food pairing, culinary fingerprinting and cuisine classification. Classification of cuisines is an interesting problem with applications for recipe recommendation and generation of novel recipes. Hitherto, cuisine classification has been attempted using ingredients of recipes as a feature. These approaches have overlooked key factors such as cooking techniques and their order which clearly form a key aspect of recipes.

We account for the loss of these information by including these details for the cuisine classification problem. As opposed to treating recipes as an itemset, we propose a methodology that views the problem as a Text Classification (TC) problem. TC refers to annotating text with one category or other based on content words and their collocations. In this article, we propose different architectures for cuisine classification on sequential datasets.

II. Related Work

Availability of detailed recipe data has evoked the interest in recipe recommendation and generation. In the past, classification of cuisines has been attempted on the basis of various factors such as time, ethnicity and place by creating six different feature stylistic sets from the data document [1]. Han Su et. al. [2] have worked on cuisine identification by using ingredients used in recipes as a basis. By identifying ingredients as features, they could provide insights on cuisine similarity. A personalised cuisine recommendation system based on user's preferences has also been proposed [3] where user's preferences are derived from their browsing activities.

Support Vector Machines [4] and several other machine learning techniques have also been implemented towards generation of a cuisine. Recently, a study on classification of cuisine on the basis of the recipe's ingredients [5] suggested a detailed relation between a recipe and its ingredients.

In this article, we propose that, beyond ingredients, even the processes and utensils involved in cooking a recipe and their order of occurrence can provide significant insights into the cuisine. We have used RecipeDB (site) [6] dataset. And to test our hypothesis, we perform classification on the dataset using several machine learning techniques, neural networks and transformers.

III. Dataset

RecipeDB was used as the source of structured data on recipes for the analysis. The dataset contains 118,071 recipes obtained from sources like AllRecipes, Epicurious Food Network, and TarlaDalal. The dataset consists of 26 cuisines as shown in Table II. Moreover, it contains an aggregation of 20280 unique ingredients, 256 unique processes and 69 unique utensils. Sample dataset of RecipeDB can be seen in Table I. Our analysis involves the following substructures of cooking recipes pertaining to traditional recipes, namely, recipes, ingredients, processes and utensils.

RecipeDB consists of ingredients, processes and utensils mined from unstructured recipe scraped from the above mentioned resources. The substructures for the recipes are mined in a sequential fashion depending upon the order in which they are used in preparing the dish. The dataset is highly sparse with a sparsity ratio of 99.50%. Out of the 20,400 distinct ingredients obtained, 11738 occur at most in one recipe such as 'lasagna noodle wheat', while 'add' appeared 1,88,004 number of times. The corresponding cumulative frequency table for the

2473-3490/20/$31.00 ©2020 IEEE
DOI 10.1109/ICDEW49219.2020.00008

TABLE I
SAMPLE DATASET FROM RECIPDB

Recipe ID	Continent	Cuisine	Recipe
2610	African	Middle Eastern	['water', 'red lentil', 'rom tomato', ...,'smooth', 'stir', 'heat']
3957	Asian	Southeast Asian	['olive oil', 'onion', 'garlic', 'ginger', ..., 'stir', 'add', 'cook ', 'season', 'garnish', 'pot']
4153	Asian	Indian Subcontinent	['coconut milk', 'milk', 'white sugar', 'basmati rice', ..., 'stir', 'cook', 'saucepan', 'bowl']
79897	Latin American	Mexican	['beef', 'chunky salsa', 'mushroom', 'garlic', ..., 'heat', 'simmer', 'serve', 'skillet']
138976	European	Deutschland	['oven buttermilk biscuit', 'onion', 'cream', ..., 'spread', 'sprinkle', 'bake', 'pan']
149191	North American	Canadian	['raisin', 'fig', 'water', 'date', 'butter', ..., 'chill', 'cut', 'bowl', 'processor', 'pan']

TABLE II
DATASET INFORMATION

Cuisine	Number of Recipes	Cuisine	Number of Recipes
Australian	5823	Japanese	2041
Belgian	1060	Korean	668
Canadian	6700	Mexican	14463
Caribbean	3026	Middle Eastern	3905
Central American	460	Northern Africa	1611
Chinese and Mongolian	5896	Rest Africa	2740
Deutschland	4323	Scandinavian	2811
Eastern European	2503	South American	7176
French	6381	Southeast Asian	1940
Greek	4185	Spanish and Portuguese	2844
Indian Subcontinent	6464	Thai	2605
Irish	2532	UK	4401
Italian	16582	US	5031

TABLE III
FREQUENCY DISTRIBUTION OF FEATURES

Number of Features	Frequency	Number of Features	Frequency
304	>1000	<2	11738
106	>5000	<3	14015
57	>10000	<4	15002
43	>15000	<5	15620
34	>20000	<6	16073
24	>25000	<7	16394
19	>30000	<8	16627
17	>35000	<10	17016
13	>40000	<15	17314
12	>45000	<20	17519

number of items shown in Table III, represents the nature of the dataset.

IV. PREPROCESSING

Before performing cuisine classification on RecipeDB data, preprocessing was implemented on structured and sequential lists of ingredients, processes and utensils. Furthermore, the digits or symbols were omitted from the items to only keep words, thereby reducing the noise in this highly sparse dataset. The preprocessing further involved tokenization followed by lemmatization of the dataset, resulting in 20,400 distinct entities.

The data is further processed to conform to the classification model requirements. Since an individual word itself doesn't impart any semantic or syntactic significance to the classification models that require quantified features as inputs, each item was translated to vectors using two techniques, namely, TF-IDF vectorization and word embedding. Depending upon the preprocessing method used, the models employed can be broadly classified into two categories: sequential models and statistical models. If the dataset is sequential, it is evident that sequential models like RNNs work better while for non sequential datasets, models like Logistic Regression, SVM, etc. perform better.

Word embeddings are essentially word representation as vectors such that semantically similar words have similar vectors whereas TF-IDF vectorization method observes the sequence of items as distinct words. Thus, TF-IDF vectors don't preserve the sequential nature of the data. Yet, we used TF-IDF technique because of its weighted function which reduces the effect of high frequency yet less meaningful words and provides a good analytical cause.

V. CLASSIFICATION

Classifying the recipes region/cuisine-wise, based on the elements involved in cooking a recipe is a major and the most important part of our analysis. The analysis treats recipes either as a sequential or as an unordered set of items. Many state-of-the-art machine learning models with TF-IDF vector inputs, such as Naive Bayes, Logistic Regression, Ensemble models along with boosting and Support Vector Machines, were tested for text classifications. Also, sequence models such as Recurrent Neural Networks and state-of-the-art NLP transformers such as BERT and RoBERTa were tested on our dataset to analyse the 'sequential nature' of the dataset. We will further discuss the implementations of these classifiers in the 'Experiments' section.

A. Naive Bayes

Naive Bayes (NB) classifier is probabilistic in nature. It is based on the supposition that all the features are independent and autonomous. NB selects the label which maximizes the posterior probability:

$$P(C_k|x) = P(C_k) * P(x|C_k)/P(x) \quad (1)$$

while the naive supposition is:

$$P(x_i|x_{i+1},\ldots,x_n,C_k) = P(X_i|C_k) \quad (2)$$

In spite of the fact that the naive supposition is false most of the time, NB gave extremely competitive results with respect to other classifiers.

B. Logistic Regression

Logistic Regression (LG) is the most used and a fundamental classifier. LG is also probabilistic in nature. It is based on the following Sigmoidal equation:

$$S(x) = 1/(1+e^{-x}) \quad (3)$$

LG treats the problem as a generalized linear regression model, which can be expressed as:

$$f(k,i) = \beta_{0,k} + \beta_{1,k}x_{1,i} + \beta_{2,k}x_{2,i} + \cdots + \beta_{M,k}x_{M,i}, \quad (4)$$

Here, for our multi-class classification problem LG is trained on a one-vs-rest scheme. Similar steps were followed in a previous research on a different dataset [5] where LG presented with the best results. This hold true in our cse as well in comparison with other baseline models.

C. Support Vector Machine

Support Vector Machines (SVMs) have been demonstrated to be having the best performance when working with textual data [4]. For classification, SVMs require translation of a multi-class classification problem to binary classification. For this, the One-vs-All approach was used. Single classifier per class was trained with the training set belonging to that class annotated as positive while the rest of the samples as negative. A strong real-valued confidence score along with a class label, by the base classifiers is required for the decision. SVM then searches for the two best-fit parallel hyperplanes which separates the two classes of data, so that they are farthest from each other.

D. Random Forest with Boosting

Random forest (RF) is a bagging decision tree approach [7]. When used as a classifier, it might not perform that well when working with a small number of features. But given that our problem is characterized with a large number of features, techniques such as RF with AdaBoost can turn out to be a good text classifier.

E. RNN (LSTM)

Recurrent Neural Networks follow a temporal or sequential connection between nodes of a layer [8]. Therefore, they are an upgrade to the conventional neural networks which consider mutual independence among the sequential inputs. Furthermore, RNNs contain an internal 'memory' and hence making them suitable for remembering previous inputs. Therefore, the characteristics of RNNs align with our problem.

We employed a state-of-the-art RNN, the Long-Short Term Memory based neural network (LSTM) [9]. LSTMs are more complex than simple RNNs as they involve a cell- and gate-like input, output and forget gate. Using these gates it controls the flow of information through the temporal dimension. It decides whether any piece of information is significant for the broader or immediate goal, or should it be removed from the memory. On account of this significant characteristic, we employed a simple 2-layer LSTM.

F. Transformer

RNNs go through words in a temporal fashion and if the sequence is long as in case of RecipeDB, the model tends to forget the crucial features of sequentially distant features. In order to overcome this limitation, attention based transformers were developed. Transformers [10] are the NLP models which are used to boost the speed of attention based models by enabling parallelization. They completely eliminate the recurrence with self attention to establish relationship between input and output, thus making them suitable for multiple language processing applications.

We employed BERT-base [11] and RoBERTa [12] models on RecipeDB. Both models perform bidirectional encoding implementing transformers after pre-training with the exception that RoBERTa is trained differently. RoBERTa was trained on longer sequences for more training steps than BERT.

VI. Experiments

For the purpose of cuisine classification we implemented different machine learning models. We tested the accuracy of the models on the RecipeDB dataset to validate the results obtained. The data was divided into 7:1:2 ratio to obtain training, validation and testing datasets respectively. Therefore out of 1,18,071 recipes training, validation and testing datasets consist of 82,650, 12,021 and 23,380 recipes respectively.

Since recipes were represented as sequences of ingredients, processes and utensils all concatenated together, long sequence were generated. The sequences were pre-processed differently for statistical models and sequential models as described in Section IV. By feeding the features obtained as input to the classifiers mentioned in Section V yielded results shown in Table IV. The corresponding code and relevant files are present in the GitHub repository: https://github.com/cosylabiiit/cuisine-classification.

Among the various statistical models that were implemented for cuisine classification, Logistic Regression performed the best, but with an accuracy of only 57.70%. These models learn the frequency of occurrence of an ingredient or process or utensil to obtain the features unique to a cuisine instead of treating the recipes as an interrelationship among these items. Furthermore, since the dataset is sparse the models couldn't fit better, leading to high bias.

Owing to the temporal relationship among the items in recipes, we observed that sequential models perform better than the statistical models. However, LSTM model gave a lower accuracy than Logistic Regression and Linear kernel SVM used in [5]. The lower accuracy is justified as the model is among the most simplistic models in the recurrent neural

TABLE IV
PERFORMANCE METRICS OF APPLIED MODELS

Dataset	Performance metric	LogReg	Naive Bayes	SVM (linear)	Random Forest	LSTM	Transformers	
							BERT	RoBERTa
RecipeDB	Accuracy	57.70	51.64	56.60	50.37	53.61	68.71	73.30
	Loss	1.51	7.14	2.97	2.32	1.65	0.21	0.10
	Precision	0.56	0.50	0.54	0.48	0.53	0.58	0.67
	Recall	0.57	0.51	0.56	0.50	0.54	0.60	0.71
	F1 Score	0.56	0.50	0.54	0.49	0.53	0.57	0.69

network class. Furthermore while comparing the LSTM model with Transformers, the sequences are treated differently i.e. LSTMs consider left to right sequence order unlike the bi-directional check in Transformers. Moreover, despite having better memory logic than vanilla RNNs, LSTMs are limited by the number of words in the sequence which further reduces the accuracy.

The limitations of LSTMs have been overcome in Transformers as explained earlier which resulted in the optimal accuracy of 73.3% and a loss of 0.10 on the RecipeDB dataset. Hence, the model is able to predict the class with least errors on minimum number of datasets among the models tested. The model presents a high average precision, recall and F1 score values representing its ability for cuisine classification.

VII. CONCLUSIONS

This article investigates different approaches for cuisine classification as a synthesis of ingredients, processes and utensils inherent to a cuisine. It also examines the effect of temporal relationships among the features to fingerprint the worldwide cuisines with the state-of-the-art RoBERTa model giving optimal results for the problem. Thus, we present a strategy to treat recipes as chains of events that are similar for a region and simultaneously contrasting from others to some extent to enable classification. Further, this articles has raised issues that can help optimise the results in different computational contexts such as recipe generation and recipe recommendation.

Apart from this, the article also raises some new research questions relating to cuisine classification. While our analysis considered for the sequential nature of recipes, the relationship among the three substructures remains unaccounted. Moreover, what features aid or hinder the classification of a recipe which could help one to uniquely distinguish between the cuisines? While maintaining the sequential nature of the recipes, redundant features were not removed. Hence, future analysis needs to identify the effect induced by these features on the classification accuracy of the models. Furthermore the imbalance among the classes affects the cuisine prediction accuracy of the classifiers. This can be reduced by ignoring the low frequency classes but would lead to a limited exploration of the world cuisines. This trade-off presents as a dilemma in this analysis.

We believe this article adds another dimension to the existing body of research on cuisine classification. This is of value for cuisine classification of unknown recipes and also aids in identifying salient features intrinsic to a cuisine.

VIII. ACKNOWLEDGEMENT

G.B. thanks the Indraprastha Institute of Information Technology (IIIT-Delhi) for providing computational facilities and support. T.S and U.U. are Research Interns in Dr. Bagler's lab (Complex Systems Laboratory) at the Center for Computational Biology and thank IIIT-Delhi for the support.

REFERENCES

[1] Y. HaCohen-Kerner, H. Beck, E. Yehudai, M. Rosenstein, and D. Mughaz, "Cuisine: Classification using stylistic feature sets and/or name-based feature sets," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 8, pp. 1644–1657, 2010.

[2] H. Su, T.-W. Lin, C.-T. Li, M.-K. Shan, and J. Chang, "Automatic recipe cuisine classification by ingredients," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*. ACM, 2014, pp. 565–570.

[3] M. Ueda, M. Takahata, and S. Nakajima, "Recipe recommendation method based on user's food preferences," in *Proceedings of the IADIS International Conference on e-Society*, 2011, pp. 591–594.

[4] J. Weston and C. Watkins, "Multi-class support vector machines," Citeseer, Tech. Rep., 1998.

[5] J. Naik and V. Polamreddi, "Cuisine classification and recipe generation," *Online: http://cs229. stanford. edu/proj2015/233 report. pdf*, 2015.

[6] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh *et al.*, "Recipedb: A resource for exploring recipes," *Available at SSRN 3482237*, 2019.

[7] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[8] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.

[9] M. Sundermeyer, R. Schlüter, and H. Ney, "Lstm neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[12] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/340778233

DESIGN AND FABRICATION OF WING OF A HIGH-PAYLOAD AERIAL VEHICLE

Article · March 2020

DOI: 10.24247/ijmperdapr202049

CITATIONS
0

READS
100

4 authors, including:



Pranjal Shiva
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Design and Fabrication of wing of a high payload Aerial vehicle View project

All content following this page was uploaded by Pranjal Shiva on 20 April 2020.

The user has requested enhancement of the downloaded file.

International Journal of Mechanical and Production Engineering Research and Development (IJMPERD)
ISSN (P): 2249–6890; ISSN (E): 2249–8001
Vol. 10, Issue 2, Apr 2020, 551–562
© TJPRC Pvt. Ltd.

TRANS STELLAR
•Journal Publications •Research Consultancy

DESIGN AND FABRICATION OF WING OF A HIGH-PAYLOAD AERIAL VEHICLE

PRANJAL SHIVA, GURU DEEP SINGH, RUPANJALI KUKAL & SANJAY KUMAR

Department of Mechanical Engineering, Delhi Technological University, New Delhi, India

ABSTRACT

The demand for an Unmanned Aerial Vehicle (UAV) is increasing by day by day due to its wide applications in diverse sectors, ranging from commercial applications to military and disaster relief missions. This paper deals with the design optimization and fabrication of the wing, which has a vital role in the lift generation. The design methodology involves a multidisciplinary system engineering approach towards the selection of pertinent factors. Wing design begins with the conceptual design and configuration selection, wherein the measurements of the wing are deduced. Balsa wood, Birch wood and aluminium were selected for the wing fabrication. A live I/O MATLAB script was utilized to calculate the dimensions of the web, flange, and related load-bearing members. The code also checks for the failure of the spar's web under shear buckling and optimizes the thickness of the web accordingly. For a better understanding of the behaviour of materials and their response to various loads, Finite Element Analysis was done on Ansys Static Structural software. Eventually, the fabrication of the designed wing was carried out with the aid of CNC router cut and the constituent parts were assembled.

KEY WORDS: Unmanned Aerial Vehicle, Finite Element Method, MATLAB, Low-Cost Prototyping, WING & Ansys

Received: Jan 12, 2020; **Accepted**: Feb 02, 2020; **Published**: Mar 12, 2020; **Paper Id.**: IJMPERDAPR202049

Original Article

INTRODUCTION

Aerostructure [1, 2] must produce safe and reliable systems with minimal weight keeping in mind the cost-effectiveness and operational conditions of the vehicle. Many of these operating conditions are harsh as they operate in extreme environments including a wide range of temperatures, pressure, and environmental loads. The airplane was the first major technology with weight as an overriding concern where majorly, all the performance parameters are affected inversely with the change in weight. The mission and range of an aircraft affect the weight variations in the system. Along with the total weight, the weight distribution also plays a major role in the designing of the control surfaces with the impact of the range, manoeuvrability and the performance of the aircraft. Aircraft wing requires structural elements like stringers, flanges, and webs that support the distribution of weight and various loads [3].

A fixed-wing UAV is an unmanned vehicle developed for long-range high endurance missions like search and rescue, mapping, surveillance, and defense. The fixed wings are responsible for the production of lift in response to forward accelerating force. The fixed wing UAV provides more efficient aerodynamics and has longer flight duration when compared to a multirotor.

This paper focuses on the application of basic structural theory[4] and knowledge of materials to design, develop and fabricate the wing of a high endurance and high payload fixed-wing UAV. The fundamental problem solved is the identification of materials, shape and dimensions of a wing of a fixed-wing UAV used for a high payload mission. The design methodology involves a multidisciplinary system engineering approach towards the

selection of various factors and performance index. The objectives of the paper are i) deduction of the measurements of the wing, ii) material selection and computational study of their response to various loading conditions, iii) optimisation of the thickness of the web by running a live I/O MATLAB script, iv) structural analysis of wing using Finite Element Method, and finally v) the fabrication and testing of the fixed wing.

DESIGN PROCESS

Conceptual Design Procedure

A similar class of UAVs was considered RMRC Air Titan [5], RMRC Anaconda [6], Sig Rascal 110 [7] to deduce the conceptual design parameters, where in payload capacity and high endurance are given precedence during selection. To ensure the sustainability of aircraft during extreme maneuvers without fracture, the design factor and the lift load factor were chosen as 1.5 and 5g respectively. Furthermore, the initial configuration was selected single-wing after narrowing down from several available configurations. Some other parameters namely, aspect ratio, wingspan, chord length were calculated. Lastly, S1223 [8], a high lift airfoil was chosen due to its suitable characteristics.

Table 1: Wing Parameters

S. No.	Parameter	Value/Range
1.	Root chord	390 mm
2.	Tip chord	277.7 mm
3.	Empty weight	< 5 kg
4.	Lift load	5g
5.	Design factor	1.5
6.	Exposed half span	1200 mm
7.	Exposed half wing area	$0.81m^2$

Preliminary Design Procedure

Empirical Relations

A variety of wing structures [9-12] were studied and the following range of parameters was assumed:

- Distance between adjacent ribs: 10-12 cm. A higher distance would not only decrease the strength of the wing but also would destroy the aerodynamic shape required for the wing. At the same time, a lower distance would increase the number of ribs, adding to unnecessary weight to the structure. Further, a constant distance between the ribs allows easy calculation and fabrication.
- Position of front spar: 18-25% of the chord length. The distance between the front spar and aerodynamic centre should be minimized to prevent torsion. Similarly, it should not be kept too close to the rear spar, as it would render the function of the front spar as futile.
- Position of rear spar: 62-65% of the chord length. The rear spar not only adds to the strength but also acts as a supporting member for the aileron (control surface used for the roll of aircraft). Thus, a too far position would diminish the thickness of the spar, resulting in it to be ineffectual.
- Position of the centre of pressure: 45% of the chord length. Although the position of the centre of pressure varies with the angle of attack, this value is suitable and can be used for calculation purposes.

Material Selection

- Balsa wood – ribs, leading-edge and skin. The material has a low density and exhibits high strength characteristics. The wood fibres can be aligned along the direction of the load to bear maximum strength.
- Birchwood – rear spar and stringers. Although it has a specific gravity higher than that of Balsa wood, the yield strength is much higher and can be used in members subjected to a higher load.
- Aluminium – Front spar. The front spar experiences the maximum force and thus needed to be a metal. Aluminium, when compared to steel, has an advantage over weight, and can also sustain the load experienced by the front spar.

Detail Design and Mathematical Model

The wings bear an enormous lift force and the structure requires to be durable. At the same time, over-building the structure would add to unnecessary weight of the aircraft. Thus, an intelligent algorithm needed to be articulated that took into account various factors, affecting the design of the wing. A MATLAB live I/O script was developed which calculates the optimum dimensions of spar for the wing. The following flow chart describes the algorithm used in the code.

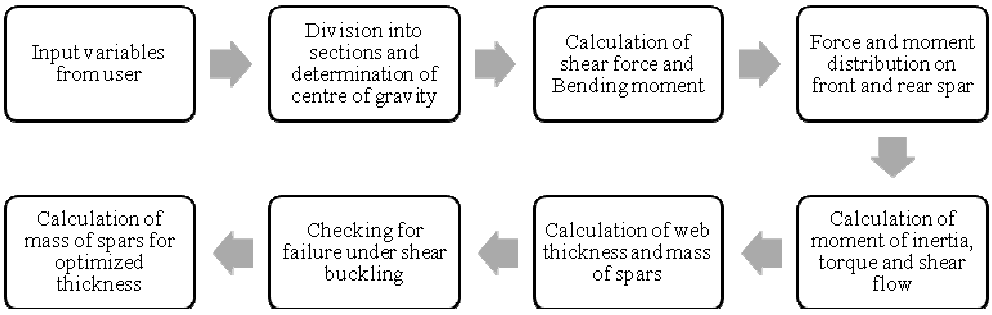


Figure 1: Algorithm for I/O MATLAB Script.

The following excerpts are included from MATLAB code:

- Parameters inputted into the code

```
prompt1='Enter the mass in kg ';
mass=input(prompt1);
prompt2='Enter the root chord in mm ';
root=input(prompt2);
```

Figure 2: MATLAB Code (i).

- Division of wing into 10 sections and calculating the centre of gravity of each trapezoid section:

```
area=0.5*(root+tip)*span;
width_each_section=span/10; %10 sections%
h=width_each_section;
l(1)=tip;
for i=2:11
    l(i)=root-((root-l(1))/span)*(span-(h*(i-1)));
    area_tr(i)=0.5*(l(i)+l(i-1))*h;
    cg(i)=h/3*((tip+2*l(i))/(tip+l(i)));
end
```

Figure 3: MATLAB Code (ii).

- Calculation of shear force and bending moment:

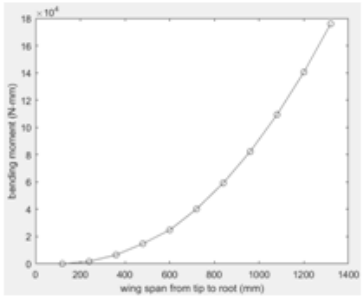
```
design_factor=1.5;
design_load=limit_load*design_factor;
design_load_half=design_load/2;
pr=design_load_half/area;

for i=2:11
    if(i==2)
        load(i)=pr*area_tr(i);
    end
    load(i)=pr*area_tr(i)+load(i-1);
end

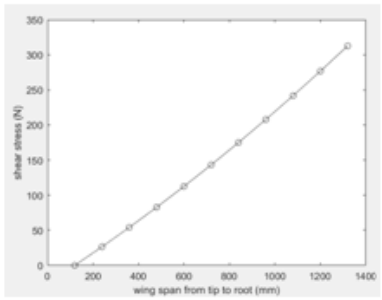
for i=2:11
    p(i)=pr*area_tr(i);
end

bm(1)=0;
bm(2)=p(2)*cg(2);
bm(3)=p(2)*cg(2)+p(3)*cg(3)+p(2)*h;
```

(A)



(B)



(C)

Figure 4: (A) MATLAB Code (iii); (B) Variation of Shear Stress along the Span; (C) Variation of Bending Moment Along the Wingspan.

- Distribution of forces on the front and rear spar:

```
    %a=location of front spar from cp
%b=location of rear spar from cp
%c=a+b
for i=1:11
    poscp(i)=cp_from_le*l(i);
    posfp(i)=fs_from_le*l(i);
    posrp(i)=rs_from_le*l(i);
    a(i)=poscp(i)-posfp(i);
    b(i)=posrp(i)-poscp(i);
    c(i)=a(i)+b(i);
end

for i=1:11
    shearfs(i)=load(i)*b(i)/c(i);
    shearrs(i)=load(i)-shearfs(i);
```

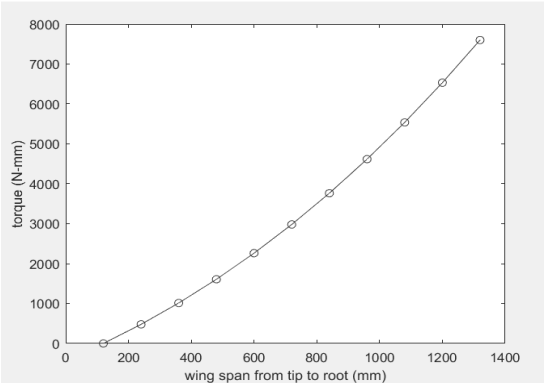
Figure 5: MATLAB Code (iv).

- Moment of inertia and torque developed

```
for i=1:11
    moifs(i)=bmfs(i)*thickfs*0.5/tensile;
    moirs(i)=bmrs(i)*thickrs*0.5/tensile;
end

torque_cum(1)=0;
for i=2:11
    areabox(i)=0.5*(thickfs+thickrs)*c(i);
    cgbox(i)=c(i)/3*((thickrs+(2*thickfs))/(thickrs+thick
    distance_cg_cp(i)=cgbox(i)-b(i);
    torque(i)=load(i)*distance_cg_cp(i);
    torque_cum(i)=torque(i);
    torque_cum(i)=torque_cum(i)+torque_cum(i-1);
    g(i)=torque_cum(i)/(2*areabox(i));
end
```

(A)



(B)

Figure 6: (A) MATLAB Code (v); (B) Variation of Torque Along the Wingspan.

- Thickness of the web and mass of spars:

```
for i=1:11
    twebfs(i)=totalfs(i)/(thickfs*shear);
    areawebfs(i)=thickfs*twebfs(i);
    moiwebfs(i)=(twebfs(i)*(thickfs)^3)/12;
    twebrs(i)=totalrs(i)/(thickrs*shear);
    areawebrs(i)=thickrs*twebrs(i);
    moiwebrs(i)=(twebrs(i)*(thickrs)^3)/12;
end
```

Figure 7: MATLAB Code (vi).

- Optimization by checking for failure under shear buckling:

```
tweb1=1;
a_new=1;
for i=2:11
    a=twebfs(i);
    j=0;
    while((a~=tweb1)&&(j<100))
    a=a_new;
    fin=q(i)/a;
    tweb1=(fin*thickfs*thickfs/(K*young))^0.5;
    a_new=tweb1;
    j=j+1;
    end
    final_web_fs(i)=a;
end
```

Figure 8: MATLAB Code (vii).

- Calculation of mass of spars for optimized thickness:

```
for i=2:11
    areanewwebfs(i)=final_web_fs(i)*thickfs;
    areanewwebrs(i)=final_web_rs(i)*thickrs;
    volnewfs(i)=(areanewwebfs(i)+aflanfs(i))*h;
    volnewrs(i)=(areanewwebrs(i)+aflanrs(i))*h;
end

volnewtotal=sum(volnewfs)+sum(volnewrs);
massnew=densityfs*volnewtotal
```

Figure 9: MATLAB Code (viii).

Table 2: Comparison of Shear Web Thickness

Rib	Front Spar Web Thickness (mm)		Rear Spar Web Thickness (mm)	
	Before Optimization	After Optimization	Before Optimization	After Optimization
1	0	0	0	0
2	0.002	0.0694	0.071	0.434
3	0.0046	0.1001	0.149	0.601
4	0.0077	0.10262	0.234	0.706
5	0.0114	0.1497	0.327	0.820
6	0.0156	0.1715	0.427	1.124
7	0.0203	0.1920	0.535	1.423
8	0.0256	0.2116	0.651	1.617
9	0.0315	0.2304	0.774	1.907
10	0.0380	0.2485	0.905	2.195
11	0.0451	0.2661	1.044	2.791

Table 3: Mechanical Parameters for Front and Rear Spar

Rib	Chord Length (mm)	Bending Moment on the Front Spar	Bending Moment on the Rear Spar	Moment of Inertia of the Front Spar	Moment of Inertia of the Rear Spar	Torque	Shear Flow	Shear Force on the Front Spar	Shear Force on the Rear Spar
1	277.77	0	0	0	0	0	0	0	0
2	288.99	664.61	938.27	32.16	15.13	530.49	0.11	14.36	16.65
3	300.22	2680.40	3784.09	129.69	61.03	1124.01	0.23	29.29	33.97
4	311.44	6099.83	8611.53	295.15	138.89	1783.00	0.35	44.79	51.94
5	322.66	10249.22	14469.49	495.93	233.38	2509.92	0.47	60.86	70.57
6	333.88	16628.93	23476.14	804.62	378.65	3307.22	0.60	77.50	89.87
7	345.11	24573.93	34692.61	1189.06	559.56	4177.33	0.74	94.70	109.82
8	356.33	34132.31	48186.79	1651.56	777.20	5122.70	0.88	112.48	130.43
9	367.55	45356.48	64032.68	2194.66	1032.78	6145.79	1.02	130.82	151.70
10	378.77	58298.87	82304.29	2820.91	1327.49	7249.05	1.17	149.74	173.64
11	390	73011.87	103075.59	3532.83	1662.51	8434.91	1.32	169.22	196.23

The final mass of the spars was deduced as 199.89 grams from the optimized thickness, higher than that from initial thickness by 80 grams. A CAD model was then designed on SolidWorks according to the dimensions obtained from the code.

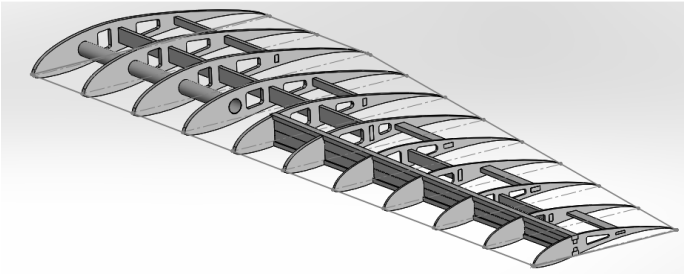


Figure 9: CAD Model of Wing.

ANALYSIS

The wing, designed in SolidWorks was analyzed on ANSYS Structural. Material properties of Aluminium 2024, Birchwood and Balsa wood were inserted from the library. The wing was then transformed into a mesh of finite elements. Simulating an actual scenario, the root rib was secured by applying boundary conditions. Further, load, calculated using the elliptical lift distribution curve was applied to the center of pressure of individual ribs. Finally, self-weight was enforced by inputting the acceleration due to gravity in the vertically downward direction.

Table 4: Force Applied on Ribs

Rib	Force (in N)
1	4.026
2	8.053
3	12.424
4	13.92
5	15.876
6	16.452
7	16.797
8	17.142
9	17.142
10	17.026
11	16.452

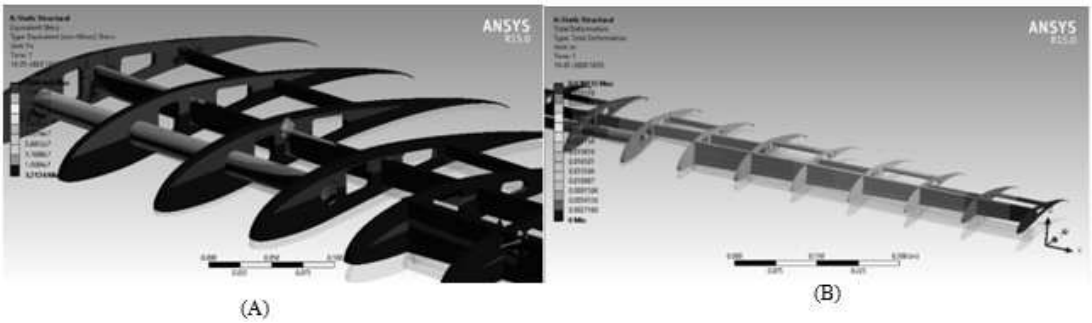


Figure 10: (A) Von-Mises Stress (B) Total Deformation Contours.

Table 5: Result of Structural Analysis

Paramter	Value
Total Deformation	3.8 cm
Maximum principal stress	197 MPa
Equivalent elastic strain	0.0128
Maximum shear stress	97 MPa
Equivalent (von-mises) stress	169 MPa
Yield strength of Aluminum	270 MPa
Safety factor	1.37

FABRICATION

The individual components of the wing were CNC router cut and were assembled over a 2-D drawing sheet. Areas bearing high load, such as the front spar, rear spar, root ribs were bonded using high strength epoxy, while others were secured using cyanoacrylate. The alignment of the wooden grains was kept parallel to the direction of the load to maximize strength. The servo motor box was installed and subsequently, the aileron was attached to the wing. The leading edge was machined to form an aerodynamic shape and finally, the wing was covered with monokote.

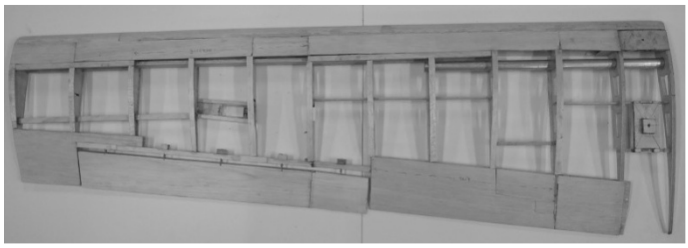


Figure 11: Assembled Structure.



Figure 12: Machining of Leading Edge.

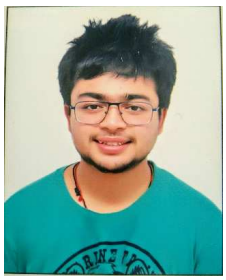
RESULTS AND DISCUSSIONS

The paper determines an effective design methodology for the structural design, optimization and fabrication of a UAV wing. The complete fabrication was carried out in a span of 11 days and the final weight of the wing was recorded as 800 grams. A safety factor of 1.37 yielded after structural analysis validates the design and the objective of the investigation is accomplished. The designed wing, thus, can be concluded to be staunch, safe and efficient for lifting heavy payloads.

REFERENCES

1. *Raymer, D. P. (1999). Aircraft design: a conceptual approach, American Institute of Aeronautics and Astronautics. Inc., Reston, VA, 21.*
2. *Sadraey, M. H. (2012). Aircraft design: A systems engineering approach. John Wiley & Sons.*
3. *Weisshaar T.A (2011). Aerospace Structures- an Introduction to Fundamental Problems Purdue University*
4. *NS, Sarath Raj, et al. "are Drones a Threat to Civil Aviation? A Retrospective on Policies of Drones in the United Arab Emirates." (2016).*
5. *Megson, T. H. G. (2016). Aircraft structures for engineering students. Butterworth-Heinemann.*
6. *Lear, D. J. (2017). Development of a Data Acquisition System for Unmanned Aerial Vehicle (UAV) System Identification.*
7. *Michailidis, M. G., Agha, et al (2017, August). A controller design framework for a NextGen circulation control based UAV. In 2017 IEEE Conference on Control Technology and Applications (CCTA) (pp. 1542-1549). IEEE*
8. *Stacul, Adrian, and Edgardo Comas. "Filtering and Acquisition of Pcm Frames Using System Generator."*
9. *Ryan, A., et al (2007, July). Decentralized control of unmanned aerial vehicle collaborative sensing missions. In 2007 American Control Conference (pp. 4672-4677). IEEE*
10. *Ma, R., et al. (2009, July). Numerical simulation of low-Reynolds-number and high-lift airfoil S1223. In Proceedings of the World Congress on Engineering (Vol. 2, pp. 1-3)*
11. *Ahmed, Md Faiyaz, and Yeole Shivraj Narayan. "Fabrication and Testing of Quadcopter Prototype for Surveillance."*
12. *Garre, P., & Arjun, G. V. (2017). Modeling and analysis of a RIBS and Spars of an airplane wing for bending and shear loads. International Journal for Research in Applied Science and Engineering, 5(2), 295-315.,*
13. *Girennavar, M., et. al (2017). Design, Analysis and Testing of Wing Spar for Optimum Weight. International Journal of Research and Scientific Innovation (IJRSI), 4(VII), 104-112.,*
14. *Ajith, V. S., et. al (2017). Study of Optimal Design of Spar Beam for the Wing of an Aircraft. International Journal of Engineering Development and Research, 5(3), 179-193.*
15. *Anton, S. R., et. al (2012). Multifunctional unmanned aerial vehicle wing spar for low-power generation and storage. Journal of Aircraft, 49(1), 292-301.*

AUTHOR'S PROFILE



Pranjal Shiva, I am currently pursuing Bachelors of Technology degree in Mechanical Engineering with specialization in Automotive Engineering from Delhi Technological University, due June 2020. My primary research areas include solid mechanics, finite element analysis, smart materials and dynamics of Unmanned Aerial Vehicles (UAVs). I am currently the project manager of an undergraduate UAV research team – Unmanned Aerial Systems (UAS-DTU) and accomplishing a project under the guidance of Indian Air Force. I have represented my college at various national and international level competitions and won many accolades. Moreover, I was recently awarded a gold medal for securing the first position in my department. I was also the founding director of Flaire Unmanned Systems Pvt. Ltd., which was later acquired by Adani Defence Systems and Technologies Limited (ADSTL) for large scale production. I am an active member of the Society of Automotive Engineers (SAE) and the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). My publications and technical design reports are as follows:

1. Technical report for design and fabrication of regular and micro sized Unmanned Aerial Vehicle for Aero Design Challenge at SAE India Southern Section.
2. Technical report for design of an autonomous launch pad at TATA Pioneer's Makerthon.
3. White paper for design of Unmanned Aerial Vehicle for Mehar Baba Prize Competition organized by Indian Air Force.
4. Recently submitted a research paper titled "Numerical Study of Fracture Parameters for Slit Specimens for Al2124 and Micro Alloyed Steel" at ICAPIE – 2019, to be published in Lecture Notes in Mechanical Engineering.
5. Recently submitted a research paper titled "Shape-Changing Airfoil and Actuated Elevons For Micro-Aerial Vehicle: UniMorph" at ICMPC – 2020.

AUTHOR'S PROFILE



Guru deep Singh, Education Details – B.Tech in Mechanical Engineering with Specialization in Automobile Engineering from Delhi Technological University (Will be Completed in 2020)

Publications –

1. “The Design and Validation of Engine Intake Manifold using Physical Experiment and CFD” published in “International Journal of Automobile Engineering Research and Development” Paper ID: IJAuERDDEC20191

2. “Design and Development of Inertia Dynamometer for FSAE Application” published in “International Journal of Advance Research and Innovation” Paper ID: IJARI-ME-19-03-110

3. "Design and Optimization of Crankshaft for Four Cylinder 4-Stroke Spark Ignition Engine using Transient Structural Analysis" published in "International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) " Paper ID: IJMPERDOCT2019106



Rupanjali Kukal is pursuing her bachelors of technology in Mechanical Engineering with specialization in Automotive engineering from Delhi Technological University (formerly Delhi College of Engineering), conducting research in the field of structural and computational field. She has in depth theoretical and practical knowledge in solid mechanics, finite element analysis and structural designing.



Sanjay Kumar, Assistant Professor at Mechanical Engineering Department, Delhi Technological University, Areas of Interest include Machine Design, Finite Element Method, Fracture Mechanics, Dynamic Fracture, Recent Publications in reputed journals are:

- Sanjay Kumar,” Comparison of Deflection and slope of Cantilever Beam with Analytical and Finite Element Method for Different Loading Conditions” International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 5, Issue 6, November 2016: ISSN: 2319-5967.
- Sanjay Kumar, Vikrant Tiwari, “Study the Effect of Pulse Shaper on Dynamic Compressive Behaviour of AL2014-T6 and AL7075-T6” International Journal of Advance Research and Innovation, Volume 4, Issue 1 (2016) 121-124 ISSN 2347 – 3258.
- Sanjay Kumar, Anoop Kumar Pandouria, Vikrant Tiwari, “Evaluation of Load and Load Point Displacement Through Three Point Bend Set Up for Dynamic Loading” International Journal of Advance Research and Innovation Volume 4, Issue 1 (2016) 114-117 ISSN 2347 – 3258.

- Pawan Kumar, Kishore P Kolhe, Sanjay Kumar," Prediction of Heat Affected Zone and Effect of Heat Input in GTA Welded Al Alloy 6061" International Journal of Engineering Science and Innovative Technology (IJESIT), Volume 4, Issue 6, November 2015: ISSN: 2319-5967

- Some conference publications:

  - Sanjay Kumar, Vikrant Tiwari "Evaluation of static fracture toughness(KIc) of Al2014-T6" Internatıonal Conference on Recent Advances ın Mechanıcal Engıneerıng (RAME2016) Department of Mechanical Engineering , Delhi Technological University Delhi-42 on 15-15 October 2016.

  - Sanjay Kumar, Anoop Kumar Pandouria and Vikrant Tiwari "Design and fabrication of Three Point Bend Set up for Dynamic loading at 5th International Symposium on Fusion of Science &Technology(ISFT-2016) "for International Conference on Advance in Science and Engineering (ICASET-2016),held at New Delhi, India on 18-22 January, 2016.

  - Sanjay Kumar, Mihir Bhambri, Udaivir Yadav, & Mayank Dhuria "Dynamic Analysis of Electromagnetic Retarder" in proceedings of International Conference Smart Technologies for Mechanical Engineering 25-26 October, 2013, DTU, Delhi, India.

View publication stats

Microchimica Acta (2020) 187: 275
https://doi.org/10.1007/s00604-020-04233-7

ORIGINAL PAPER

Electrochemical biosensor for the epithelial cancer biomarker EpCAM based on reduced graphene oxide modified with nanostructured titanium dioxide

Owais Jalil[1] · Chandra Mouli Pandey[1] · Devendra Kumar[1]

Received: 30 August 2019 / Accepted: 23 March 2020 / Published online: 18 April 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

An electrochemical immunosensor has been fabricated for the early determination of epithelial cell adhesion molecules (EpCAM, tumor biomarker) antigen using reduced graphene oxide (rGO) modified with nanostructured titanium dioxide (TiO_2). The hydrothermally synthesized rGO@TiO_2 nanocomposite has been electrophoretically deposited on indium tin oxide (ITO) coated glass substrate, and the deposition was confirmed using various spectroscopic, microscopic, and electrochemical techniques. The fabricated rGO@TiO_2/ITO electrode shows improved electron transfer kinetics with an electron transfer rate constant of 1.93×10^{-7} cm·s^{-1}. Furthermore, the rGO@TiO_2/ITO electrodes were used for the covalent immobilization of monoclonal EpCAM antibodies. Electrochemical determination of the EpCAM cancer biomarker is achieved using differential pulse voltammetry by scanning the potential from − 0.4 to 0.8 V with an amplitude of 50 mV. The rGO@TiO_2-based biosensor shows high sensitivity (3.24 μA·mL·ng^{-1}·cm^{-2}), wide detection range (0.01 ng·mL^{-1} to 60 ng·mL^{-1}), and low detection limit (0.0065 ng·mL^{-1}, S/N = 3). The fabricated biosensor is highly stable and regenerable and has been successfully applied to the determination of EpCAM in spiked human serum samples.

Keywords Biomarker · Cancer · Epithelial cell adhesion molecules · Electrophoretic deposition · Electrochemical biosensor · Graphene oxide · Titanium dioxide

Introduction

The early detection and quantification of cancer biomarkers provide an efficient pathway to observe tumor progression and relate them to specific cancer types. In this context, epithelial cell adhesion molecules (EpCAM) have been considered a prognostic tumor biomarker for cancer diagnosis [1–3]. EpCAM is a glycosylated transmembrane protein, which is expressed on the surface of circulating tumor cells, and the overexpression of EpCAM has generally been found in the carcinomas of the breast, stomach, colon, rectum, prostate, biliary tract, pancreas, etc. [4, 5]. For the detection of EpCAM, techniques such as cytometry, enzyme-linked immunosorbent assays, and polymerase chain reaction have been widely used [6, 7]. However, these methods are usually expensive, complicated, time-consuming, and even require highly skilled human resources [8]. Therefore, there is an urgency for the development of an effective alternative method for early, sensitive, affordable, and reliable determination of EpCAM.

Some immunosensor based on electrochemical [9], fluorescence [10, 11], surface plasmon resonance [12], and colorimetric techniques have been reported for the detection of EpCAM. Shiddiky et al. have successfully functionalized 2D graphene oxide (GO) nanosheets with quantum dots and antibodies and demonstrated that these bionanoconjugates are highly sensitive and specific for the electrochemical detection of EpCAM [13]. A

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00604-020-04233-7) contains supplementary material, which is available to authorized users.

✉ Chandra Mouli Pandey
cmp.npl@gmail.com

✉ Devendra Kumar
dkumar@dce.ac.in

[1] Department of Applied Chemistry, Delhi Technological University, Delhi 110042, India

fluorescence-based biosensor using graphene quantum dots and molybdenum disulfide nanosheets have also been fabricated for rapid and sensitive detection of EpCAM [11]. Despite these achievements, still, there is a need for the fabrication of a sensitive, selective, and cost-effective biosensor that can be used for early and reliable detection of EpCAM. Nowadays, electrochemical biosensors are being widely used because of their simplicity, high sensitivity, smooth operation, accuracy, and selectivity [14]. For the fabrication of efficient electrochemical biosensor, the selection of effective transducer surfaces that bind the recognition molecule and providing faster access to the analyte is required [15]. In this context, GO has attracted tremendous interest in the fabrication of low-cost electrode materials because of its unique physical properties, such as high specific surface area, carrier mobility, electrical conductivity, flexibility, and optical transparency [16]. The oxygen-containing functional groups in GO make it strongly hydrophilic, thus facilitating the chemical functionalization that helps in the conjugation of nanoparticles and biomolecules [17–19]. Anchoring of metal oxide (ZnO, TiO_2, CuO) nanoparticles on the surface of GO with homogenous distribution exhibited improved electronic properties and served as an excellent platform for various applications [20]. Among them, TiO_2 nanoparticles have attracted considerable interest due to their superior properties, such as their large specific surface area, high uniformity, and excellent electrochemical properties. The porous nature of GO permits the penetration and adsorption of TiO_2 into the surface of GO, resulting in large surface area, better electron mobility along with more active sites for immobilization of biomolecules on the electrode surface [21, 22].

Keeping this in view, efforts have been made to synthesize rGO@TiO_2 nanocomposite using hydrothermal techniques where the GO was reduced using L-cysteine. Furthermore, rGO@TiO_2 was used as a sensing platform for the determination of EpCAM by covalently immobilizing anti-EpCAM antibodies. The response characteristics of this immunosensor pertaining to the interaction of the anti-EpCAM/rGO@TiO_2/ITO electrode with EpCAM antigen has been carried out using differential pulse voltammetry (DPV). To the best of our knowledge, the present study is the first report on the fabrication of rGO@TiO_2 nanocomposite-based transducer platform for the electrochemical determination of EpCAM cancer biomarker.

Materials and methods

All chemicals, materials, and instrumentation details are discussed in electronic supporting materials (ESM).

Synthesis of GO and TiO_2 nanoparticles

Modified Hummer's method was used for the synthesis of GO from graphite. In brief, 0.450 g of graphite powder was added to a mixture of concentrated H_2SO_4:H_3PO_4 (9:1) with constant stirring. After 10 min, 9 g of $KMnO_4$ was added to the mixture and the stirring was continued for another 12 h at 35 °C. After 12 h, the reaction was stopped and the reactants are allowed to cool at room temperature. To remove the excess of $KMnO_4$, 1.2 mL of H_2O_2 was dropped slowly and stirred for 15 min and the temperature was maintained below 20 °C. The suspension was washed with HCl (20 mL) and DI water (60 mL) and was centrifuged at 5000 rpm for 10 min. The obtained semi-solid material was vacuum dried and the solid brown powder of GO was thus obtained. The synthesized GO was dispersed in DI water (0.1 $mg{\cdot}mL^{-1}$) and sonicated for further use [23]. For the synthesis of TiO_2 nanoparticles, 15 mL of tetrabutyl titanate (TBOT) was slowly added into 1500 mL of DI water under vigorous stirring. After stirring for 3 h and aging for another 24 h, the formed gel precipitate was filtered and washed with DI water. The obtained precipitate was dried at 80 °C for 10 h and was finely powered and calcined at 500 °C for 3 h [24].

Synthesis of rGO@TiO_2 nanocomposite

For the synthesis of the rGO@TiO_2 nanocomposite, GO (10 mg) and TBOT (1.5 mmol) were added into diethylene glycol (30 mL), at constant stirring for 12 h at room temperature. The formed mixture was then added drop-wise to 180 mL of acetone, followed by the slow addition of DI water. The mixture was again stirred for 2 h that resulted in the formation of GO@TiO_2 nanocomposite that was centrifuged (15,000 rpm) and washed with water/ethanol [24]. For the reduction of the GO, the synthesized GO@TiO_2 and L-cysteine (1.5 mmol) were added into DI water (80 mL) and transferred to 100 mL autoclave for hydrothermal reaction (120 °C; 12 h). The obtained product was centrifuged (15,000 rpm), and the precipitate was collected and washed with water/ethanol (Fig. 1a) [25].

Electrophoretic deposition of rGO and rGO@TiO_2 nanocomposite

The electrophoretic deposition (EPD) of the synthesized rGO and rGO@TiO_2 nanocomposite onto hydrolyzed ITO electrode was carried out using DC power supply (Genetix, Model GX300C) using Pt as the counter electrode. For the EPD, the colloidal solution (5 mL) of rGO@TiO_2 nanocomposite (0.5 $mg{\cdot}dL^{-1}$) was diluted with 10mL ethanol which was then electrophoretically deposited onto hydrolyzed ITO electrode at 12 V for 30 s. Magnesium nitrate (10^{-4} M) was added to the above colloidal solution to obtain an adequate

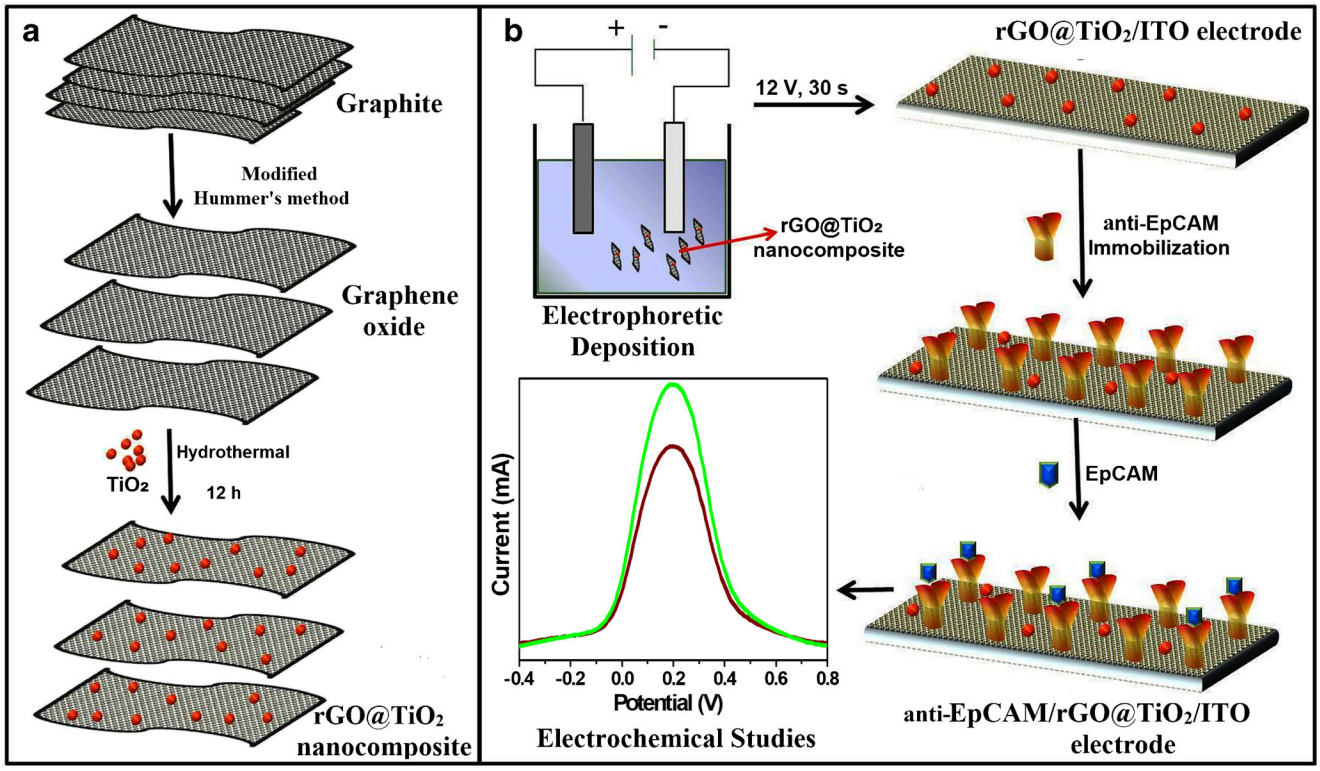


Fig. 1 Schematic illustration for the synthesis and electrophoretic deposition of rGO@TiO_2 on ITO coated glass substrate (**a**) and subsequent biosensor fabrication (**b**)

positive surface charge on the rGO@TiO_2 nanocomposite, as the Mg^{2+} ions get absorbed onto the rGO sheets. After the EPD, the electrode was taken out from the solution, washed with DI water, and dried [26].

Fabrication of rGO@TiO_2/ITO-based immunosensor

For the fabrication of immunosensor, rGO@TiO_2/ITO electrodes were treated with EDC (2 mM; 10 μL) and NHS (5 mM; 10 μL) for 1 h at room temperature (27 °C), to activate the –COOH functional groups. Furthermore, the modified electrodes were incubated with 20 μL (50 μg mL^{-1}) of monoclonal antibodies (anti-EpCAM) in PBS containing Tween-20 (0.05%). To avoid non-specific antigen binding, the anti-EpCAM/rGO@TiO_2/ITO electrodes were treated with BSA (1%) for 2 h to block the exposed binding sites of the electrode, followed by washing with PBS (3 times). The schematic for the fabrication of the immunosensor has been shown in Fig. 1b.

Choice of materials

A novel electrochemical biosensor based on rGO@TiO_2 nanocomposite has been proposed for the determination of EpCAM cancer biomarker. Nanocomposites based on GO@TiO_2 have been previously reported, but the application was limited mainly to lithium-ion batteries, photocatalysis, or energy-related fields [27]. Very few reports for utilizing GO@TiO_2 nanocomposite in sensing applications are available still, more research is needed to explore this highly desirable nanocomposite material in the biosensing field. Furthermore, the extremely large surface area and remarkable conducting properties of GO along with a high isoelectric point of TiO_2 may provide a new opportunity to utilize this rGO@TiO_2 nanocomposite for designing efficient biosensors. The modified electrode (rGO@TiO_2/ITO electrodes) can effectively enhance the electrochemical performance along with more active sites for the immobilization of biomolecules.

Results and discussion

The TEM images of GO sheet display a crumbled and re-stacked morphology (Fig. 2a) with a large surface area (few micrometers in the diameter). This large surface area provides an advantage for the growth of TiO_2 nanoparticles on the GO surface. Figure 2b shows the TEM micrograph of rGO@TiO_2 nanocomposite, where spherically polycrystalline TiO_2 nanoparticles (30–40 nm) are evenly distributed on the GO sheet. So, it can be revealed that GO acts as an excellent platform for entrapping TiO_2 nanoparticles. The formation of GO, rGO, and rGO@TiO_2 nanocomposite has also been confirmed

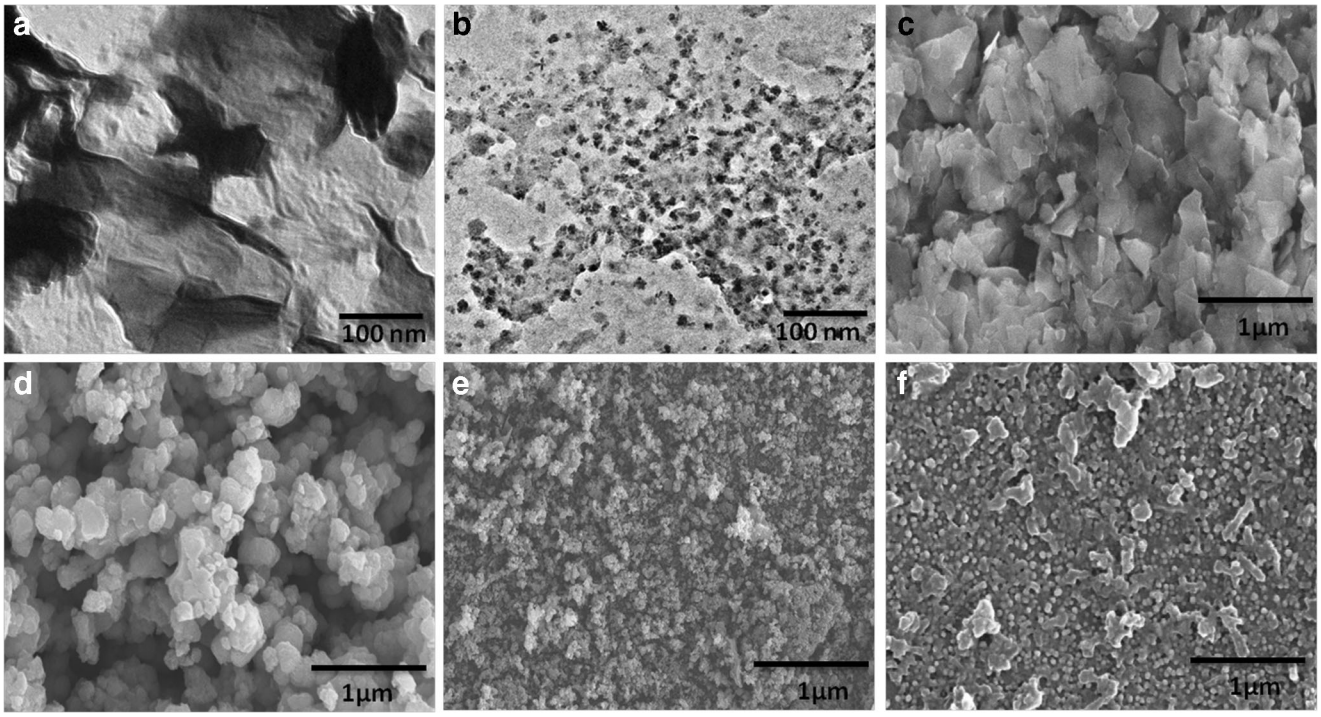


Fig. 2 Transmission electron microscopic image of GO (**a**) and rGO@TiO_2 nanocomposite (**b**). Scanning electron microscopic images of rGO/ITO electrode (**c**), TiO_2/ITO electrode (**d**), rGO@TiO_2/ITO electrode (**e**), and anti-EpCAM/rGO@TiO_2/ITO electrode (**f**)

using X-ray diffraction pattern (Fig. S1A and B), UV-vis spectrum (Fig. S2A and B), TGA (Fig. S3A), and FTIR (Fig. S3B and C) (respective text and figures are given in the Electronic Supporting Material).

The SEM image of the rGO/ITO electrode (Fig. 2c) shows a large sheet-like structure with thin curtains like morphology, which is due to the delaminating of graphite during the oxidation process. Figure 2d shows the SEM micrograph for the TiO_2 nanoparticles, having a spherical shape with good dispersion and less agglomeration. Furthermore, the surface morphology of GO changes after modification with TiO_2 and shows the surface of rGO covered with spherical TiO_2 nanoparticles (Fig. 2e). After the immobilization of anti-EpCAM onto the surface of rGO@TiO_2/ITO electrode, the granular morphology becomes globular due to covalent attachment of antibodies (Fig. 2f).

Electrochemical characterization

Electrochemical studies of the rGO/ITO electrode, rGO@TiO_2/ITO electrode, and anti-EpCAM/rGO@TiO_2/ITO bioelectrode have been performed in PBS containing 5 mM of $[Fe(CN)_6]^{3-/4-}$ using electrochemical impedance spectroscopy (EIS) in the frequency range 0.01–10^5 Hz. The electrochemical impedance spectra (Nyquist plots) mainly include a semicircle portion at higher frequencies, followed by a linear part at lower frequencies corresponding to electron transfer–limited process and a diffusion-limited electron transfer process, respectively [28]. The semicircle diameter gives the value of charge transfer resistance (R_{ct}) and can be derived numerically by fitting an equivalent circuit based on the modified Randles and Ershler model [29]. The characteristic Nyquist plots for the fabricated electrodes are shown in Fig. 3a along with the fitted Randles equivalent circuit (Fig. 3a; inset). The circuit consists of an electron transfer resistance (R_{ct}), constant phase element (C_{PE}), solution resistance (R_s), and Warburg impedance (W) [30]. The EIS curve of rGO/ITO electrode shows a R_{ct} value of 1.35 kΩ (curve (ii)), which is attributed to good conductivity of GO that facilitates fast transfer of electrons. However, for the rGO@TiO_2/ITO electrode (curve (i)), there is a decrease in the R_{ct} value (1.12 kΩ). This decrease in R_{ct} can be assigned to the synergistic effects of GO sheets and TiO_2 that enhance the electrochemical performance of rGO@TiO_2/ITO electrode. Furthermore, after immobilization of anti-EpCAM onto rGO@TiO_2/ITO electrode (curve (iii)) and blocking of exposed electrode site using BSA, there is an increase in the R_{ct} value (1.68 kΩ). This increase in R_{ct} can be due to the non-conducting nature of antibody that hinders the charge transfer pathway. When the target antigen was incubated to anti-EpCAM/rGO@TiO_2/ITO electrode, an antibody-antigen complex has been formed that creates a barrier for the electrochemical transduction process, thereby hindering the access of the redox probe to the electrode surface and thus increasing the R_{ct} (2.53 kΩ) [31]. In the EIS spectra, the reaction occurring at the electrode is governed

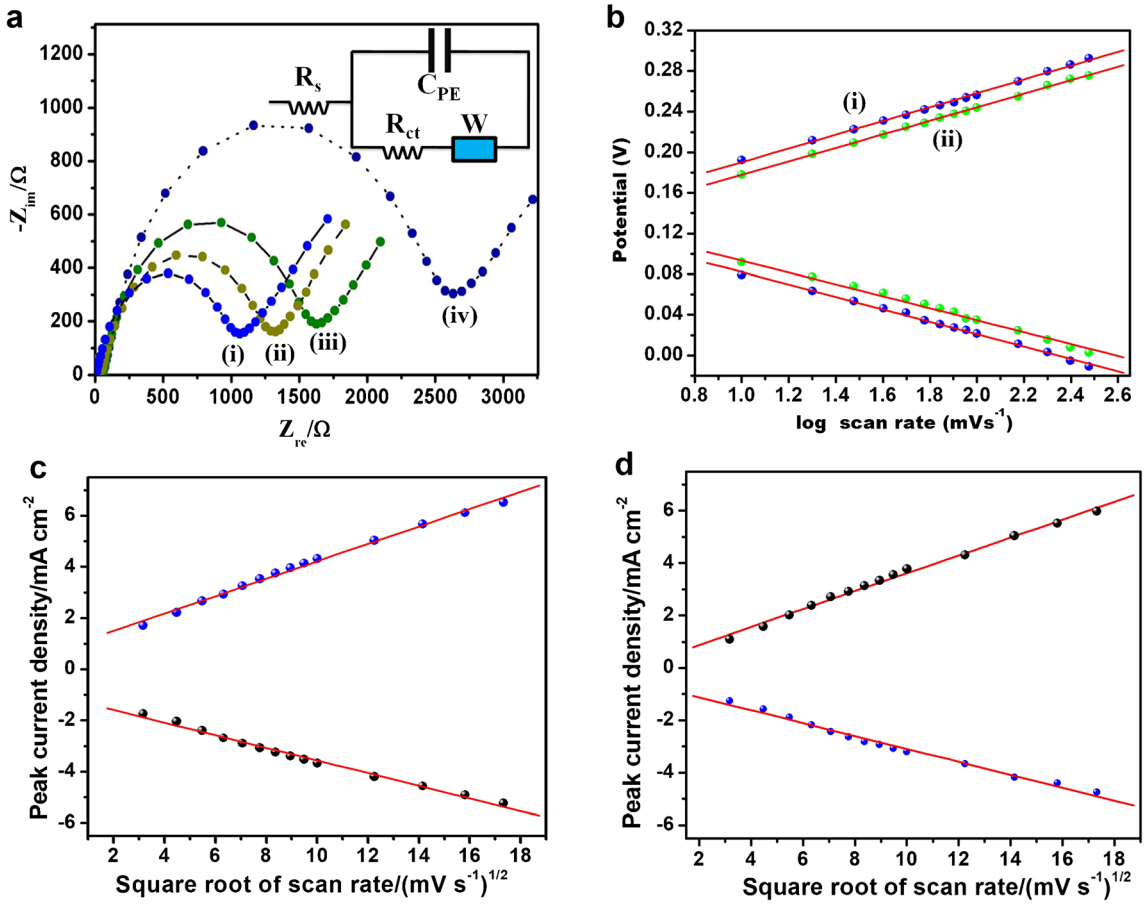


Fig. 3 (**a**) Nyquist diagram for the Faradic impedance for (i) rGO@TiO_2/ITO electrode, (ii) rGO/ITO electrode, (iii) anti-EpCAM/rGO@TiO_2/ITO electrode, and (iv) EpCAM immobilized on anti-EpCAM/rGO@TiO_2/ITO electrode measured in PBS solution (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$, in the frequency range from 10^5 to 10^{-2} Hz. (**b**) The plots of potential with log of scan rate for (i) rGO@TiO_2/ITO electrode and (ii) rGO/ITO electrode. The plot for I_{pa}, I_{pc}, vs. square root of scan rate (10–300 mV) for the rGO@TiO_2/ITO electrode and (**c**) rGO/ITO electrode (**d**) in PBS solution (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$

by the electron–transfer kinetics. Thus, the exchange current per geometric unit area ($i_o = nRT/R_{ct}F$) for the rGO/ITO electrode and rGO@TiO_2/ITO electrode is found to be 7.60×10^{-5} A·cm^{-2} and 9.32×10^{-5} A·cm^{-2}, respectively. The apparent electron transfer rate constant (k_{app}) for the modified electrodes is also calculated using the equation: $k_{app} = RT/n^2F^2AR_{ct}C$; where F is the Faraday constant, n denotes the number of electrons transferred, R is the gas constant, T is the temperature in Kelvin, A is the area of the electrode, and C is the concentration of the $[Fe(CN)_6]^{3-/4-}$. The values of k_{app} are found to be 1.57×10^{-7} cm·s^{-1} and 1.93×10^{-7} cm·s^{-1} for the rGO/ITO electrode and rGO@TiO_2/ITO electrode, respectively. It has been observed that the peak current density of cyclic voltammetry (CV) is proportional to the square root of the scan rate, revealing the linear dependence of peak currents to the scan rate. This indicates that the electron transfer from rGO/ITO and rGO@TiO_2/ITO electrode is a surface-controlled process [32]. When the scan rate (υ) was increased from 10 to 300 mV·s^{-1}, the oxidation peak drifts to a more positive value and the reduction peak to more negative values. This shows that both the peak potentials (E_p and E_c) are linearly dependent on the log υ (Fig. 3b) and yields two straight lines (Eqs. 1, 2, 3, and 4) with anodic and cathodic slopes of $2.303RT/(1-\alpha)nF$ and $-2.303RT/\alpha nF$, respectively.

$$E_p(V)[rGO@TiO_2/ITO] = 0.124(V) + 0.067(V) \times Log[scanrate(mV/s)]; R^2 = 0.998 \quad (1)$$

$$E_c(V)\ [rGO@TiO_2/ITO] = 0.143(V) - 0.060\ (V) \times Log\ [scan\ rate\ (mV/s)]; R^2 = 0.994 \quad (2)$$

$$E_p(V)\ [rGO/ITO] = 0.115\ (V) + 0.067\ (V) \times \text{Log}\ [\text{scan rate}\ (mV/s)];\ R^2 = 0.996 \quad (3)$$

$$E_c(V)\ [rGO/ITO] = 0.151(V) - 0.058\ (V) \times \text{Log}\ [\text{scan rate}\ (mV/s)];\ R^2 = 0.987 \quad (4)$$

Based on the plot and the above equation, the values of α are estimated to be 0.59 for rGO/ITO electrode and 0.61 for rGO@TiO_2/ITO electrode. The effective surface area (A) was calculated using the linear slope of the anodic and cathodic peak currents density (I_{pa} and I_{pc}) on the square root of scan rate (Fig. 3c, d; Eqs. 5, 6, 7, and 8), and the Randle Sevick equation: $Ip = (2.69 \times 10^5) n^{3/2} A C D^{1/2}\ \upsilon^{1/2}$. Where Ip is the peak current, C is the molar concentration of $[Fe(CN)_6]^{3-/4-}$ ($mol \cdot cm^{-3}$), n is the electrons per molecule oxidized or reduced, D is the diffusion coefficient, and υ is the scan rate. The effective surface area (A) of the rGO@TiO_2/ITO electrode ($D = 1.31 \times 10^{-5}\ cm^2 \cdot s^{-1}$) is found to be 0.075 cm^2. The A calculated for rGO/ITO electrode and ITO electrode is found to be 0.086 cm^2 ($D = 0.89 \times 10^{-5}\ cm^2 \cdot s^{-1}$) and 0.10 cm^2 ($D = 0.53 \times 10^{-5}\ cm^2 \cdot s^{-1}$), respectively.

$$I_{pa}\ (\mu A)\ [rGO@TiO_2/ITO] = 631.4\ \mu A + 367.0\ \mu A\ (s \cdot mV^{-1}) \times \text{scan rate}\ (mV \cdot s^{-1});\ R^2 = 0.999 \quad (5)$$

$$I_{pc}(\mu A)[rGO@TiO_2/ITO] = -110.2\mu A - 246.8\mu A(s \cdot mV^{-1}) \times \text{scanrate}(mV \cdot s^{-1});\ R^2 = 0.996 \quad (6)$$

$$I_{pa}(\mu A)[rGO/ITO] = 63.1\mu A - 345.1\mu A(s \cdot mV^{-1}) \times \text{scanrate}(mV \cdot s^{-1});\ R^2 = 0.998 \quad (7)$$

$$I_{pc}(\mu A)[rGO/ITO] = -626.8\mu A - -241.7\mu A(s \cdot mV^{-1}) \times \text{scanrate}(mV \cdot s^{-1});\ R^2 = 0.987 \quad (8)$$

Optimization of biosensing parameters

The concentration of anti-EpCAM required for binding onto the surface of rGO@TiO_2/ITO electrode using EDC-NHS chemistry was also optimized using DPV and is found to be 50 $\mu g \cdot mL^{-1}$ (Fig. S4A). The maximum binding time of EpCAM (1 $ng \cdot mL^{-1}$) antigen with the anti-EpCAM/rGO@TiO_2/ITO electrode has been achieved in 10 min (Fig. S4B).

Electrochemical biosensing studies

DPV studies were carried out for ITO electrode, rGO/ITO electrode, rGO@TiO_2/ITO electrode, and anti-EpCAM/rGO@TiO_2/ITO electrode to study the effect of electrode modification at a various step (the respective data and figures are given in the Electronic Supporting Materials (Fig. S5)). To examine the efficacy of the biosensor, electrochemical studies of the anti-EpCAM/rGO@TiO_2/ITO electrode were carried out as a function of EpCAM concentration (0.01–60 $ng \cdot mL^{-1}$) in PBS containing 5 mM $[Fe(CN)_6]^{3-/4-}$ using DPV. A continuous decrease in the anodic peak current has been observed when the anti-EpCAM/rGO@TiO_2/ITO electrode was treated with an increasing concentration of EpCAM (Fig. 4a). This decrease in the response current is due to the interaction of EpCAM with anti-EpCAM at the transducer surface that results in the formation of an insulating layer (antigen-antibody complex) on the electrode surface that retards the electron transfer from the bulk solution to the electrode. Figure 4b shows the calibration plot for the change in the magnitude of the peak current of anti-EpCAM/rGO@TiO_2/ITO electrode as a function of EpCAM concentration (0.01–60 $ng \cdot mL^{-1}$) and follows Eq. (9):

$$I_p = 0.025(\mu A) - 0.814(\mu A \cdot mL \cdot ng^{-1}) \times \text{EpCAM concentration}(ng \cdot mL^{-1}) \quad (9)$$

The sensitivity (slope/surface area) calculated for anti-EpCAM/rGO@TiO_2/ITO electrode is 3.24 $\mu A \cdot mL \cdot ng^{-1} \cdot cm^{-2}$ with a correlation coefficient of 0.997. The limit of detection (LOD) calculated (3σ/sensitivity), for the fabricated biosensor is 0.0065 $ng \cdot mL^{-1}$ or 240 fM ($S/N = 3$), where σ is the standard deviation of the blank electrode. For all the sensing experiments, anti-EpCAM/rGO@TiO_2/ITO electrode with PBS solution was taken as a control signal, and the imprecision of the data (Fig. 4b) is found to be < 3%. The biosensing performance of the fabricated anti-EpCAM/rGO@TiO_2/ITO electrode has been compared with other methods for EpCAM detection (Table 1). The fabricated biosensor achieved a LOD of 0.0065 $ng \cdot mL^{-1}$ that is comparable with the other reported biosensors [11–13, 33]. The enhanced electrochemical performance of anti-EpCAM/rGO@TiO_2/ITO electrode can be attributed to the good conductivity of rGO@TiO_2 that improves the local topographic interactions with the anti-EpCAM [34]. The fabricated biosensor was also compared with other reported TiO_2-based biosensors for cancer biomarker determination, and it has been observed that the proposed method shows significant results (Table S1). To determine the affinity of EpCAM toward anti-EpCAM/

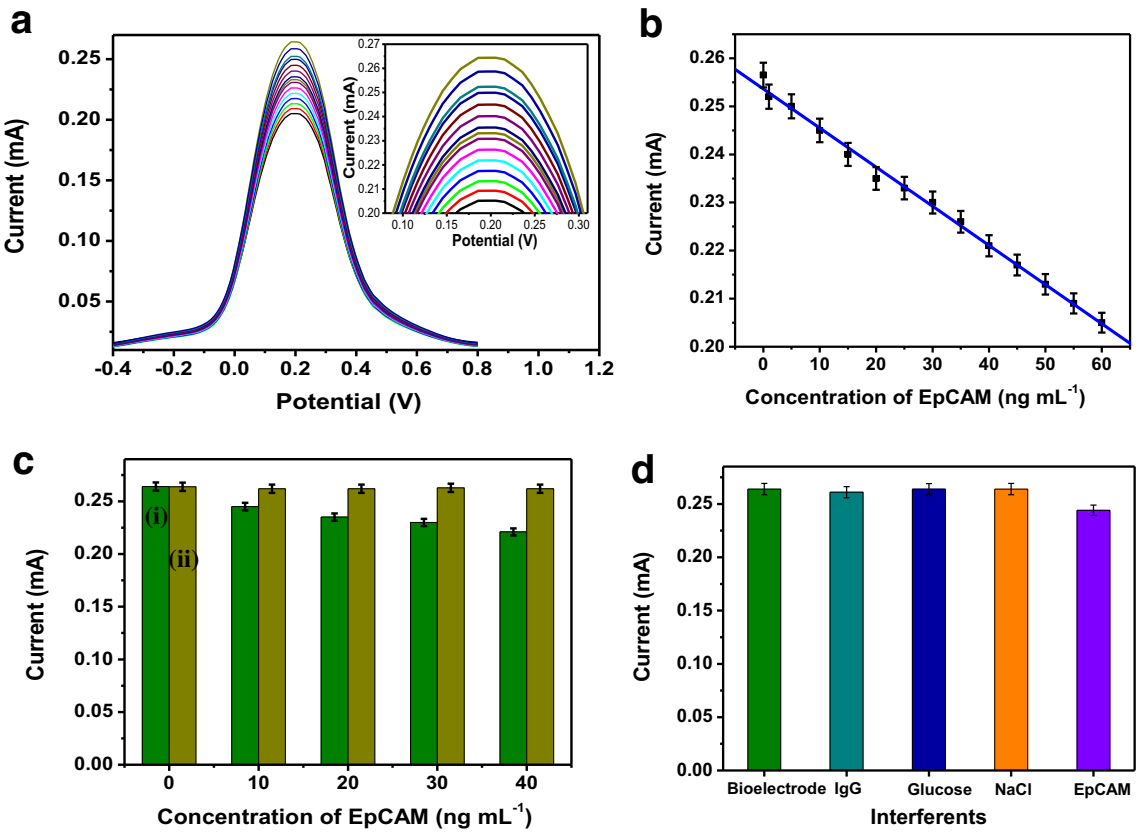


Fig. 4 Electrochemical response of the anti-EpCAM/rGO@TiO_2/ITO electrode (**a**) as a function of EpCAM concentration (from top to bottom, 0 to 60 ng·mL^{-1}) using DPV; inset is the magnified view of oxidation peak current. (**b**) Calibration plot between the magnitudes of current recorded and EpCAM concentration (0.01–60 ng·mL^{-1}). Error bars represent the standard deviations of repeated determinations (n = 3). (**c**) Control experiment showing the DPV response of (i) anti-EpCAM/rGO@TiO_2/ITO electrode and (ii) rGO@TiO_2/ITO electrode against EpCAM. (**d**) Specificity testing of anti-EpCAM/rGO@TiO_2/ITO-based biosensor with glucose and IgG proteins as control groups. The measurements were conducted by scanning the potential from − 0.4 to 0.8 V with an amplitude of 50 mV in PBS solution (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$. Error bars: SD, n = 3

rGO@TiO_2/ITO electrode, the association constant (K_a) has been calculated using the Hanes–Wolf plot [35] and the higher is the value of K_a, the stronger is the affinity of EpCAM toward anti-EpCAM/rGO@TiO_2/ITO electrode. The K_a estimated (plot between substrate concentrations versus substrate concentration/current) is found to be 4.86×10^{-15} M^{-1}.

Table 1 Biosensing characteristics of anti-EpCAM/rGO@TiO_2/ITO electrode along with those reported in the literature

S. no.	Immobilization matrix	Biomolecule	Detection technique	Linear detection range	Detection limit	Labeling	Sensitivity	Ref.
1.	GQDs-MoS_2	Aptamer	Fluorescence	3–54 nM	450 pM	Yes	–	[11]
2.	LC-SPDP/Au	Antibody	CV and Fluorescence	1×10^5–1×10^8 cells mL^{-1}	1×10^5 cells mL^{-1}	Yes	–	[12]
3.	Rubpy core-shell Silica	Antibody	Fluorescence and FCM	–	–	Yes	–	[33]
4.	GRs-QDs	Antibody	ASV	100 fg·mL^{-1} - 7.5 μg·mL^{-1}	100 fg·mL^{-1}	Yes	–	[13]
5.	AgNPs-PVA	Antibody	EM	2–2000 pg·mL^{-1}	0.8 pg·mL^{-1}	No	–	[36]
6.	MCH/Hp1/AuE	Aptamer	SWV	0.1–20 ng·mL^{-1}	20 pg·mL^{-1}	Yes	–	[37]
7.	rGO-TiO_2/ITO	Antibody	DPV	0.01–60 ng·mL^{-1}	6.5 pg·mL^{-1} or 240 fM	No	3.24 μA·mL·ng^{-1} ·cm^{-2}	Present work

GQDs, graphene quantum dots; *MoS_2*, molybdenum disulphide; *LC-SPDP*, succinimidyl 6-(3-[2-pyridyldithio]-propionamido) hexanoate; *Au*, gold; *Rubpy*, tris(bipyridine)ruthenium(II)chloride; *GRs*, graphene oxide nanosheets; *QDs*, quantum dots; *AgNPs-PVA*, silver nanoparticles covered with polyvinyl alcohol; *MCH/Hp1/AuE*, 6-mercapto-1-hexanol/hairpin probe 1/gold electrode; *FCM*, flow cytometer; *ASV*, anodic stripping voltammetry; *EM*, electrochemical microfluidic; *SWV*, square wave voltammetry

Electrochemical studies were also conducted to evaluate the current response of bare rGO@TiO_2/ITO electrode toward different concentrations of EpCAM (Fig. 4c). The obtained results indicate that there is no significant change in the peak current, revealing no false signal is generated, and the electrochemical response is completely due to the interaction between EpCAM and anti-EpCAM/rGO@TiO_2/ITO electrode.

Stability and regenerability of the biosensor

The stability and reusability of the biosensor have also been tested. It was observed that after 35 days, the current response of the immunosensor remains 93.7% of the initial value (Fig. S6A). Furthermore, the bioelectrode can be regenerated and reused, for at least 5 times with a relative standard deviation of 6.6% (Fig. S6B).

Specificity and serum sample analysis

The specificity of the fabricated immunosensor for EpCAM detection was carried out by incubating the anti-EpCAM/rGO@TiO_2/ITO electrode with glucose, immunoglobulin G (IgG), sodium chloride (NaCl), and EpCAM with the same concentration (15 ng·mL^{-1}). These analytes are typical interfering species that exist in human plasma. The bar diagram (Fig. 4d) clearly shows that no false signal has been generated when the electrode was treated with glucose, IgG, and NaCl showing good specificity of the biosensor for EpCAM detection.

To validate the performance of the fabricated biosensor, the electrochemical response studies of the anti-EpCAM/rGO@TiO_2/ITO electrode with spiked human serum samples at 5, 10, 15, and 20 ng·mL^{-1} were performed (Table S2). The spiked serum samples have been prepared by adding 20 μL of EpCAM into 100 μL of serum. The observed results exhibited acceptable recoveries (98–101%) with relative standard deviation of 1.8–4.7% (n = 3) of EpCAM in serum samples. This suggests that the fabricated immunosensor can be an alternative technique for the determination of EpCAM in human serum.

Conclusion

A highly sensitive electrochemical biosensor has been fabricated for the determination of cancer biomarker using rGO@TiO_2 nanocomposite. The environment-friendly route was used for the synthesis of the nanocomposite, which shows excellent electrochemical properties, selectivity and performs as an appropriate sensing layer. The GO-based substrate provides an excellent platform for the growth of TiO_2 nanoparticles and with covalent conjugation of anti-EpCAM, it shows a much lower detection limit. This fabricated biosensor was validated with spiked serum samples and it shows a promising result for the efficient detection of cancer biomarkers in the serum sample. It is believed that the fabricated platform has the potential to be used for the monitoring of other cancer biomarkers, which would be transformative in the diagnosis of cancer. Besides this, more efforts are needed to develop straight forward methods for the determination of the EpCAM profile for circulating tumor cells.

Acknowledgements C.M. Pandey acknowledges the Department of Science and Technology (DST), New Delhi, India, for DST-INSPIRE Faculty award (Grant No. DST/INSPIRE/04/2015/000932).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Gorges TM, Tinhofer I, Drosch M, Röse L, Zollner TM, Krahn T, von Ahsen O (2012) Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition. BMC Cancer 12:178
2. Gostner JM, Fong D, Wrulich OA, Lehne F, Zitt M, Hermann M, Krobitsch S, Martowicz A, Gastl G, Spizzo G (2011) Effects of EpCAM overexpression on human breast cancer cell lines. BMC Cancer 11:45
3. Treerattrakoon K, Chanthima W, Apiwat C, Dharakul T, Bamrungsap S (2017) Oriented conjugation of antibodies against the epithelial cell adhesion molecule on fluorescently doped silica nanoparticles for flow-cytometric determination and in vivo imaging of EpCAM, a biomarker for colorectal cancer. Microchim Acta 184:1941–1950
4. Patriarca C, Macchi RM, Marschner AK, Mellstedt H (2012) Epithelial cell adhesion molecule expression (CD326) in cancer: a short review. Cancer Treat Rev 38:68–75
5. Yoon HJ, Kozminsky M, Nagrath S (2014) Emerging role of nanomaterials in circulating tumor cell isolation and analysis. ACS Nano 8:1995–2017
6. Ntouroupi T, Ashraf S, McGregor S, Turney B, Seppo A, Kim Y, Wang X, Kilpatrick M, Tsipouras P, Tafas T (2008) Detection of circulating tumour cells in peripheral blood with an automated scanning fluorescence microscope. Br J Cancer 99:789–795
7. Zhao S, Yang H, Zhang M, Zhang D, Liu Y, Liu Y, Song Y, Zhang X, Li H, Ma W (2013) Circulating tumor cells (CTCs) detected by triple-marker EpCAM, CK19, and hMAM RT-PCR and their relation to clinical outcome in metastatic breast cancer patients. Cell Biochem Biophys 65:263–273
8. Lambrechts AC, Bosma AJ, Klaver S, Top B, Perebolte L, Van't Veer L, Rodenhuis S (1999) Comparison of immunocytochemistry, reverse transcriptase polymerase chain reaction, and nucleic acid sequence–based amplification for the detection of circulating breast cancer cells. Breast Cancer Res Treat 56:219–231
9. Shen H, Yang J, Chen Z, Chen X, Wang L, Hu J, Ji F, Xie G, Feng W (2016) A novel label-free and reusable electrochemical cytosensor for highly sensitive detection and specific collection of CTCs. Biosens Bioelectron 81:495–502
10. Li Z, Wang Y, Wang J, Tang Z, Pounds JG, Lin Y (2010) Rapid and sensitive detection of protein biomarker using a portable

fluorescence biosensor based on quantum dots and a lateral flow test strip. Anal Chem 82:7008–7014
11. Shi J, Lyu J, Tian F, Yang M (2017) A fluorescence turn-on biosensor based on graphene quantum dots (GQDs) and molybdenum disulfide (MoS2) nanosheets for epithelial cell adhesion molecule (EpCAM) detection. Biosens Bioelectron 93:182–188
12. Arya SK, Wang KY, Wong CC, Rahman ARA (2013) Anti-EpCAM modified LC-SPDP monolayer on gold microelectrode based electrochemical biosensor for MCF-7 cells detection. Biosens Bioelectron 41:446–451
13. Shiddiky MJA, Rauf S, Kithva PH, Trau M (2012) Graphene/quantum dot bionanoconjugates as signal amplifiers in stripping voltammetric detection of EpCAM biomarkers. Biosens Bioelectron 35:251–257
14. Chikkaveeraiah BV, Bhirde AA, Morgan NY, Eden HS, Chen X (2012) Electrochemical immunosensors for detection of cancer protein biomarkers. ACS Nano 6:6546–6561
15. Soni A, Pandey CM, Solanki S, Kotnala RK, Sumana G (2018) Electrochemical genosensor based on template assisted synthesized polyaniline nanotubes for chronic myelogenous leukemia detection. Talanta 187:379–389
16. Zhu Y, Murali S, Cai W, Li X, Suk JW, Potts JR, Ruoff RS (2010) Graphene and graphene oxide: synthesis, properties, and applications. Adv Mater 22:3906–3924
17. Hussein KH, Abdelhamid HN, Zou X, Woo H-M (2019) Ultrasonicated graphene oxide enhances bone and skin wound regeneration. Mater Sci Eng C 94:484–492
18. Hallaj R, Haghighi N (2017) Photoelectrochemical amperometric sensing of cyanide using a glassy carbon electrode modified with graphene oxide and titanium dioxide nanoparticles. Microchim Acta 184:3581–3590
19. Sanghavi BJ, Wolfbeis OS, Hirsch T, Swami NS (2015) Nanomaterial-based electrochemical sensing of neurological drugs and neurotransmitters. Microchim Acta 182:1–41
20. Ashour RM, Abdelhamid HN, Abdel-Magied AF, Abdel-Khalek AA, Ali MM, Uheida A, Muhammed M, Zou X, Dutta J (2017) Rare earth ions adsorption onto graphene oxide nanosheets. Solvent Extr Ion Exch 35:91–103
21. Jang HD, Kim SK, Chang H, Roh K-M, Choi J-W, Huang J (2012) A glucose biosensor based on TiO_2–graphene composite. Biosens Bioelectron 38:184–188
22. Josephine DSR, Babu KJ, Gnana Kumar G, Sethuraman K (2017) Titanium dioxide anchored graphene oxide nanosheets for highly selective voltammetric sensing of dopamine. Microchim Acta 184:781–790
23. Shahriary L, Athawale AA (2014) Graphene oxide synthesized by using modified hummers approach. Int J Renew Energy Environ Eng 2:58–63
24. Perera SD, Mariano RG, Vu K, Nour N, Seitz O, Chabal Y, Balkus KJ (2012) Hydrothermal synthesis of graphene-TiO_2 nanotube composites with enhanced photocatalytic activity. ACS Catal 2:949–956
25. Pandey CM, Tiwari I, Singh VN, Sood K, Sumana G, Malhotra BD (2017) Highly sensitive electrochemical immunosensor based on graphene-wrapped copper oxide-cysteine hierarchical structure for detection of pathogenic bacteria. Sens Actuators B Chem 238:1060–1069
26. Srivastava S, Kumar V, Ali MA, Solanki PR, Srivastava A, Sumana G, Saxena PS, Joshi AG, Malhotra BD (2013) Electrophoretically deposited reduced graphene oxide platform for food toxin detection. Nanoscale 5:3043–3051
27. Fan W, Lai Q, Zhang Q, Wang Y (2011) Nanocomposites of TiO_2 and reduced graphene oxide as efficient photocatalysts for hydrogen evolution. J Phys Chem C 115:10694–10701
28. Randviir EP, Banks CE (2013) Electrochemical impedance spectroscopy: an overview of bioanalytical applications. Anal Methods 5:1098–1115
29. Randles JEB (1947) Kinetics of rapid electrode reactions. Discuss Faraday Soc 1:11–19
30. Jorcin J-B, Orazem ME, Pébère N, Tribollet B (2006) CPE analysis by local electrochemical impedance spectroscopy. Electrochim Acta 51:1473–1479
31. Xie Y, Chen A, Du D, Lin Y (2011) Graphene-based immunosensor for electrochemical quantification of phosphorylated p53 (S15). Anal Chim Acta 699:44–48
32. Pandey CM, Sumana G, Tiwari I (2014) Copper oxide assisted cysteine hierarchical structures for immunosensor application. Appl Phys Lett 105:103706
33. Tao L, Zhang K, Sun Y, Jin B, Zhang Z, Yang K (2012) Anti-epithelial cell adhesion molecule monoclonal antibody conjugated fluorescent nanoparticle biosensor for sensitive detection of colon cancer cells. Biosens Bioelectron 35:186–192
34. Dowaidar M, Abdelhamid HN, Hällbrink M, Zou X, Langel Ü (2017) Graphene oxide nanosheets in complex with cell penetrating peptides for oligonucleotides delivery. BBA-Gen Subjects 1861:2334–2341
35. Augustine S, Joshi AG, Yadav BK, Mehta A, Kumar P, Renugopalakrishanan V, Malhotra BD (2018) An emerging nanostructured molybdenum trioxide-based biocompatible sensor platform for breast cancer biomarker detection. MRS Commun 8:668–679
36. Bravo K, Ortega FG, Messina GA, Sanz MI, Fernández-Baldo MA, Raba J (2017) Integrated bio-affinity nano-platform into a microfluidic immunosensor based on monoclonal bispecific trifunctional antibodies for the electrochemical determination of epithelial cancer biomarker. Clin Chim Acta 464:64–71
37. Chen Q, Hu W, Shang B, Wei J, Chen L, Guo X, Ran F, Chen W, Ding X, Xu Y (2018) Ultrasensitive amperometric aptasensor for the epithelial cell adhesion molecule by using target-driven toehold-mediated DNA recycling amplification. Microchim Acta 185:202

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)

FPGA Simulation of Fingertip Digit Recognition Using CNN

MD Shahbaz Khan
Department of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
mdshahbaz_bt2k16@dtu.ac.in

Niharika
Department of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
niharika_bt2k16@dtu.ac.in

Priya Yadav
Department of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
priyayadav_bt2k16@dtu.ac.in

Rishabh Verma
Department of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
rishabhverma823@gmail.com

Dr Indu Sreedevi
Department of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
s.indu@dtu.ac.in

***Abstract*—The paper proposes a novel framework for fingertip-writing digit recognition using Convolutional Neural Network (CNN) and its hardware implementation using Verilog. In the above-mentioned framework, an image of a handwritten digit was fed using SystemVerilog. Using SystemVerilog interfacing with C language, the image was read from a file and its corresponding RGB value was stored in a text file. Using Python, the text file was read, and the digit was recognized using CNN. CNN model was trained on MNIST dataset of 50000 images of handwritten digits with the accuracy rate of 98.4% for predicting an unseen digit. The predicted digit was displayed using SystemVerilog(Xilinx ISE). The proposed algorithm is fast in processing as it uses CNN for model training. Existing works in this field uses image classification tree to recognise digits[12], whose accuracy is 95.8% which is much less than the proposed algorithm in this paper.**

Keywords— CNN, fingertip-based digit recognition, FPGA simulation, image processing.

I. INTRODUCTION

Computer vision finds its wide application in day-to-day life. Character recognition is an important task of computer vision algorithm. With the progress in algorithms, computational efficiency is increasing manifold thus making human-computer interface smooth, faster and more precise. Advancement in machine learning algorithms has also developed few computational efficient computer vision algorithms and one such is Convolutional Neural Network (CNN). Fingertip based gesture system plays a vital role in the life of physically disabled people. Using fingertip gestures, one can communicate in the vacuum without any medium of writing. Such systems should be computationally faster with higher precision. The paper proposes a noble methodology of fingertip based digit recognition. Using camera input, video of the movement of a fingertip is recorded using SystemVerilog and C++ language. On applying the image processing algorithms in C++ language, the gesture of digit is extracted from video into one frame. That frame is further applied to a trained model for digit recognition. The digit predicted by the CNN based algorithm is displayed on seven Segment display, connected to the FPGA board.

II. LITERATURE LAYOUT

Several algorithms have come up to make fingertip-based digit recognition faster. However, accuracy for recognising digits is still low. Algorithms like [1], has an accuracy of 95% only, as compared to ours of 98.4%. Some of the algorithms [2], [3] uses image processing and decision tree based digit recognition, which is computationally less efficient with lower accuracy. In contrary to this, our algorithm uses FPGA implementation of CNN which makes it much faster as compared to the algorithms running on CPUs or GPUs.

Paper like [4] uses R-CNN based algorithm for fingertip tracking. However, as per the result of VOT2016 challenge, CSRDCF tracking algorithm out throws the other tracking algorithms. It has an efficiency of 87% for tracking an object in continuous frames at an fps of 50. Along with CSRDCF based fingertip tracking, the paper proposes CNN based handwritten digit recognition system. CNN model for digit recognition is trained on 50,000 handwritten images and been tested on 10,000 images which yielded an accuracy of 98.4%. The paper deals in FPGA simulation of digit recognition using CNN. As per the white paper published by Microsoft [5], training of any neural network on FPGA is much faster than its training on CPUs and GPUs.

III. PROBLEM STATEMENT

The paper is aimed to find the solution for the fingertip digit recognition system. The intent is to implement the CNN based recognition system on FPGA board to increase its computational efficiency. Earlier work in this field [6] was carried out using Python language on CPUs, which are not as fast as FPGAs. The proposed algorithm also enhances the accuracy of digit prediction as it uses CNN for handwritten digit recognition. Works like [7], [1] deals in the decision tree and SVM based classifier for handwritten digit recognition, whose accuracy is not as high as CNN based algorithm. Moreover, on implementing the algorithm on FPGA, training speed of model also increases manifold.

IV. EXTRACTING FOREGROUND

The very step in the field of fingertip tracking is the removal of background from each frame. Background subtraction is an important image pre-processing technique to extract the region of interest. This preprocessing of the image is mainly done to subtract the foreground from the background. For the abovementioned task, background segmentation algorithm (based on Gaussian Mixture) was used. This algorithm is based on two research papers [8],[9] given in the year 2004 and 2006 respectively. For each pixel, the algorithm picks the Gaussian distributions by itself. It makes the algorithm flexible for diurnal changes. Moreover,

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-7281-5475-6/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

it proves to efficient for shadow detection and its removal from the foreground. This background subtraction algorithm makes the whole framework independent of diurnal changes and shadow and hence enhances its efficiency.

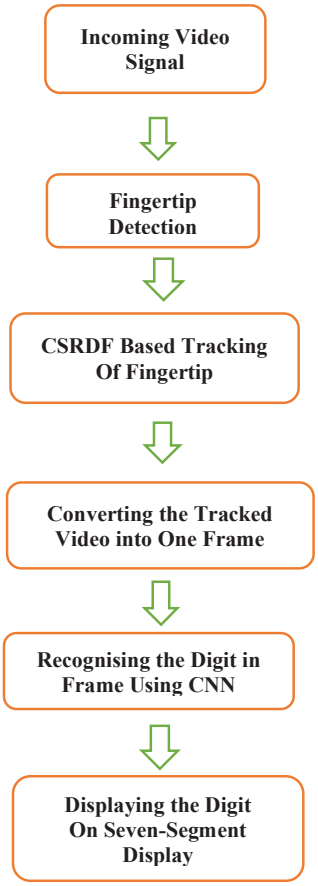


Fig. 1. Flowchart of the proposed algorithm

V. FINGERTIP DETECTION

Detecting a fingertip in an RGB video is a tough and challenging task as the dimension of the fingertip is small compared to the whole frame. However, recent CNN-based algorithms have come up to detect fingertip precisely and accurately. The paper implements an algorithm [4] which segments the hand region, followed by localization of its centroid. From the segmented region, the fingertip is localized with the help of an entropy function, termed as distance-weighted curvature entropy function.

In the above-mentioned entropy function, the distance of each contour points from the centroid is calculated and stored in a 1-D matrix. Then the weighted distance is calculated by multiplying the matrix with the distance of curvature points from the centre of the hand. Using contours detecting algorithm as proposed by [10], contours of the hand are detected. Let p be the contours of the hand of length k. On sampling the contours into f samples, the sampling interval is given by,

$$\Delta p = \frac{k}{f} \tag{1}$$

The turning angle θ, is defined as the change in direction of the tangent of two points on contour. While curvature c is the change in tangent direction of points on moving along the curve and is given by,

$$c(p) \approx \frac{\theta}{\Delta p} \tag{2}$$

Hence the curvature entropy u which is locally proportional to $c(p)$ (curvature of the contour) is approximated by,

$$u\big(c(p)\big)\theta - \cos\big(\Delta p.c(p)\big) \tag{3}$$

The other factor influencing entropy function is the distance between contour points and the centroid (x', y') which is denoted by $s(p)$. The complete function is denoted by $\chi(p)$ and is defined as the following,

$$\chi(p) = u\big(c(p)\big).s(p)^{\tau} \tag{4}$$

Where τ is the weighted distance as defined earlier.

It is known that the region of the fingertip will have maximum value for curvature entropy along with the maximum distance from the centroid. Thus, the entropy function $\chi(p)$ will have its maxima at the fingertip coordinates, which may be expressed as,

$$x_o.y_o = p(arg\ max\ \chi(p)) \tag{5}$$

Where (x_o, y_o) are the coordinates of the fingertip.

VI. FINGERTIP TRACKING USING CSRDCF

As discussed in [11], CSRDCF is a tracking algorithm which provides learning with higher efficiency and better integration of filter updates and tracking. On detecting the fingertip in a frame, the tracking algorithm(CSRDCF) starts its execution. In this algorithm, spatial reliability map is used which helps to adjust filter support to enhance the accuracy of tracking. Similarly, reliability scores are calculated to check the quality of filters and thus helps in evaluating coefficients in localization.

The paper proposes a novel framework of object tracking based on discriminative correlation filtering followed by channel and spatial reliability scores[12]. It is a proximity-based algorithm which learns a set of filters to represent the object and uses them to find the highest correlated point in the adjacency of tracking an object from the previous frame.

A. Spatially Correlation Filters

With the help of the following approach, we comprehend the importance of learning correlation filters. For a given set of 'n' channels with features f = $\{f_k\}_{k=1:n}$ and target templates h = $\{h_k\}_{k=1:n}$, and the position of the object is determined by locating the highest frequency of ĝ(h),

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k \tag{6}$$

where $\star$ is circular convolution operator & ĝ(h) is correlation response.

Similarly, to find optimum correlation filter h, we have to minimize

$$\mathcal{E}(h) = \|\hat{g}(h) - g\|^2 + \lambda \|h\|^2 \tag{7}$$

where g is the expected output.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

To find the efficiency of tracking, we transformed it into Frequency Domain as

$$\mathcal{E}(h) = \| \sum_{k=1}^{n} diag(\hat{f}_k)\hat{h}'_k - \hat{g}' \|^2 + \lambda \|h\|^2 \quad (8)$$

And on minimizing the above equation we get,

$$\hat{h}_k = (diag(\hat{f}_k)\hat{g}') \Theta^{-1} (\sum_{k=1}^{n} diag(\hat{f}_k)\hat{f}'_k + \lambda) \quad (9)$$

where Θ-1 is element-wise division operator and the above result can be used to design correlation filter trackers.

On considering each channel to be independent, we get the cost function as

$$\mathcal{E}(h) = \sum_{k=1}^{n} \|f_k \star h_k - g\|^2 + \lambda \|h\|^2 \quad (10)$$

To scale, we introduce a factor depending upon the discriminative power of feature channels which is termed as channel weights $w = \{\hat{w}_k\}_{k=1:n}$ and whose correlation response is given by,

$$\hat{g}(h) = \sum_{k=1}^{n} f_k \star h_k . \hat{w}_k \quad (11)$$

B. Correlation Filter Learning

As mentioned above, it has been assumed that the correlation filter is independent of channels, hence it allows to consider the case of one filter only. The constraints are formalized as following

$$h_c - m \Theta h \equiv 0 \quad (12)$$

where h_c is defined as the dual variable with the above constraints.

On using Lagrangian augmentation [9], we find the following outcome

$$s\mathcal{A}(\hat{h}_c, h, \hat{I}|m) = \|diag(\hat{f}_k)\hat{h}'_k - \hat{g}\|^2 + \lambda \|h_m\|^2$$

$$+ [\hat{I}^H (\hat{h}_c - \hat{h}_m) + \hat{I}^H (\hat{h}_c - \hat{h}_m)'] + \mu \|\hat{h}_c - \hat{h}_m\|^2 \quad (13)$$

Here in the above equation, we have introduced a new term μ which is the constraint penalty value and its value is updated by the following equation.

$$\mu^{i+1} = \beta\mu^{i} \quad (14)$$

C. Spatial Reliability Map

On locating the target, the region of interest is extracted from the frame and is used for updating the filter. The constrained filter uses spatial reliability map mp, which is the collection of pixels of training model matching with that of the testing mode. To determine m_p, we have designed the following algorithm.

While the algorithm is getting executed for tracking, foreground, and background of the object is stored in colour histogram c. Let $y_i = [y_i^c, y_i^x]$ is the observation i.e., it denotes the colour of y_i^c along with the position y_i^x for each pixel and m_p = random variable indicating the unknown foreground/background. For y_i to be observed on output, its joint probability was found as follows

$$p(y_i) = \sum_{j=0}^{1} p(y_i | m_i = j) p(m_i = j)$$

$$= \sum_{j=0}^{1} p(y_i^c | m_i = j) p(y_i^x | m_i = j) p(m_i = j) \quad (15)$$

TABLE I. TRACKING ALGORITHM

CSRDCF
Prerequisites Image Ii, the position of an object in previous frame pos_{i-1} , scale k_{i-1}, filter h_{i-1}, image histograms c_{i-1}, channel reliability w_{i-1}. Determination and approximation of scale:- 1. Locating target at pos_i : finding the highest frequency in a correlation between filter (h_{i-1}) and extracted features (f) at pos_{i-1} position and adjusted by reliability scores (w). 2. Calculation of $\hat{w}$(det) (detection validation) with the help of per-channel responses. 3. Scale k_i was calculated, with the help of pos_i. 4. Colour histograms for foreground as well as the background were calculated and updated using $\hat{w}$(det) (detection validation). 5. Reliability map m_p was calculated using the above-mentioned relation. 6. Filter $\hat{h}$ was calculated and updated using m_p. 7. Using h, learning channel reliability was determined. 8. Calculate channel reliability from learning channel reliability & detection reliability. 9. Each filter hi was updated using the relation $h_i = (1-\eta)h_{i-1} + \eta\hat{h}$. 10. Channel reliability w_i was updated using the expression $w_i = (1-\eta)w_{i-1} + \eta\hat{w}$.

VII. DIGIT RECOGNITION USING CNN

After tracking the fingertip movement in consecutive frames, a single frame is constructed from it as shown in figure 4. The image thus formed is further used to recognise the numeral character present in it. For recognising the handwritten digit in the frame, we have deployed CNN based algorithm. As proposed in [13], CNN is a computationally faster algorithm as compared to other multi-layer perceptron. CNN works on 2D convolution on the input pixel values. CNN based algorithm implemented in our paper for digit recognition task can be divided into the following subtasks: -

A. Convolution Layer

The very first layer of CNN is the convolution layer. An RGB image consists of three channels i.e., Red, Green and Blue. However, convolution layer can only work on a single channel at a time. Thus, 3 channel RGB image was converted into one channel by taking an average of it. Hence, in this layer, we have a 2-D matrix of 28 x 28 pixels. We picked three filters randomly of 5x5 from the image. Convolution of all the three filters with the image was performed resulting in three 2-D matrices of 28 x 28 pixels. The whole task could be summed up in a table: -

TABLE II. CONVOLUTION LAYER ALGORITHM

CONVOLUTION LAYER ALGORITHM
Step 1: RGB image was converted into a single channel image. Step 2:- Three filters of 5 x 5 pixel was picked from the image randomly. Step 3:- Each filter was multiplied by the corresponding pixel of 28 x 28 image.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

CONVOLUTION LAYER ALGORITHM
Step 4:- All the elements of the resultant 5 x 5 matrix was summed and divided by 25.
Step 5:- Filter was moved and centred at the next pixel of image and step 2 to 4 was repeated for each filter separately.

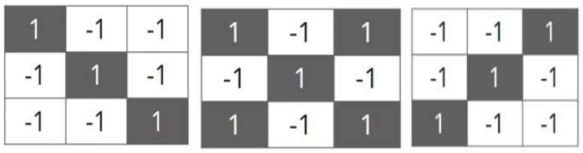


Fig. 2. Three Filters used in Convolution.

B. Activation Layer

The resultant matrices of convolution layer may or may not consists of negative values. However, a negative value as the pixel of an image holds no meaning. To remove negative values from the pixel matrix of filtered images, we applied an activation function. An activation function only lets the positive values to pass through it. There are many activation functions as such sigmoid function, Rectified Linear Unit (ReLU) function, tanh function etc.

In our algorithm, we applied ReLU as an activation function for the filtered images. A ReLU only activates a node if its value is greater than 0 and deactivates the negative node. A ReLU layer can be mathematically defined as follows: -

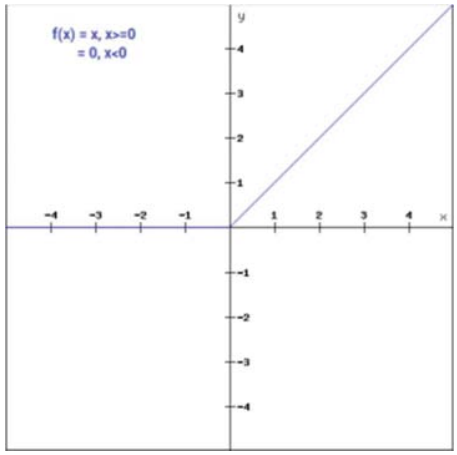


Fig. 3. ReLU Activation Function.

C. Max Pool Layer

After passing the image matrices from ReLU layer, the dimension of the matrix is very high i.e., 24 x 24 that makes up to 576 elements. To reduce the size of the image matrix, we created a window of 2 x 2 size and put it at the beginning of the image pixel matrix. The maximum value of that window was picked and replaced in the image matrix. The window was slide over the whole matrix, thus reducing its size to its half.

After pooling layer, the whole process starting from convolution to pooling was repeated thrice. This resulted in 4 x 4 matrices of three channels. All the elements of three channels were arranged in a single vector/list in a row-major form. The vector thus formed after training, was unique for each digit and were 10 in number.

Unknown handwritten digits had to undergo all the above-mentioned steps, which yielded a 1-D vector for them. The resultant vector was compared with all the 10 vectors that were formed during the training of the model. The unknown digit was predicted on the highest resemblance of its vector with the precalculated 10 vectors.

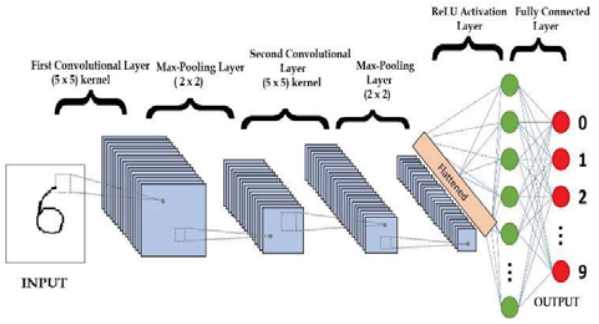


Fig. 4. Different Layers of CNN.

VIII. DISPLAY ON SEVEN-SEGMENT

After the fingertip drawn digit is recognized by CNN algorithm, it is displayed on seven-segment display board. This is task is performed using the modules created in SystemVerilog. Steps performed for displaying the digit are as follows :

TABLE III. DISPLAYING THE RECOGNISED DIGIT

STEPS FOR DISPLAYING THE PREDICTED DIGIT
Step 1: Read the predicted digit using 'readmemb()' function
Step 2: Assigned the data read to a register named 'digits'.
Step 3:Converted the binary data to the anode and cathode connection values using 'case()'.
Step 4: Produced the data if data lies in the range 0-9 else turned off the 7-segment display.

IX. RESULT

The above-proposed algorithm was trained on MNIST datasets containing 60,000 images of handwritten digits. Dataset was divided into training and testing data in the ratio 5:1. The CNN model was trained with the following parameters :-

- Number of epochs = 17
- Learning rate = 0.003
- Number of hidden layers = 2
- Optimizer used = Stochastic Gradient Descent

On testing the trained model on 10,000 images, the accuracy of the algorithm was 98.4% which is much higher than [7] which was 95%.



Fig. 5. Fingertip Tracking

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

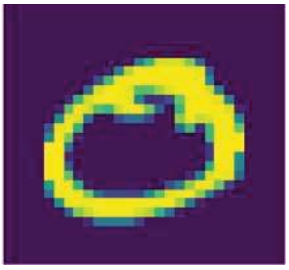


Fig. 6. Fingertip Digit fed into CNN algorithm

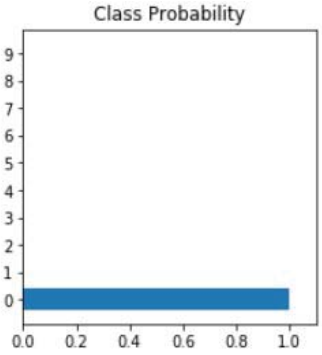


Fig. 7. Digit predicted by the CNN algorithm

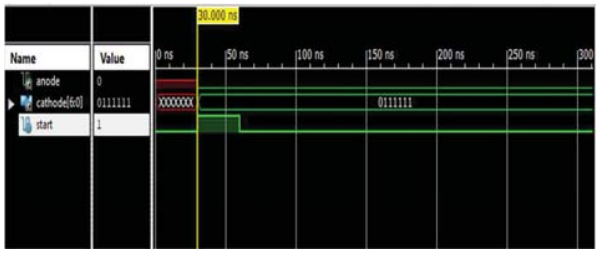


Fig. 8. Digit '0' displayed on Seven-Segment.

```
Time resolution is 1 ps
Simulator is doing circuit initialization process.
Finished circuit initialization process.
Digit predicted is (binary form) :
0000
Anode value of Seven Segment :
0
Cathode value of Seven Segment :
0111111
ISim>
```

Console | Compilation Log | Breakpoints

Fig. 9. SystemVerilog output of the predicted digit '0'.

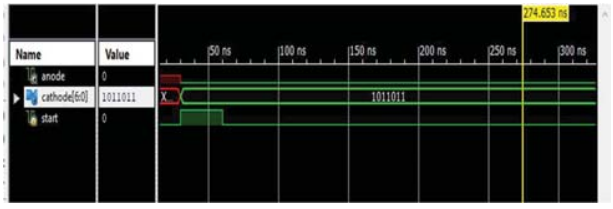


Fig. 10. Digit '2' displayed on Seven-Segment.

```
Time resolution is 1 ps
Simulator is doing circuit initialization process.
Finished circuit initialization process.
Digit predicted is (binary form) :
0010
Anode value of Seven Segment :
0
Cathode value of Seven Segment :
1011011
ISim>
```

Console | Compilation Log | Breakpoints

Fig. 11. SystemVerilog output of the predicted digit '2'.

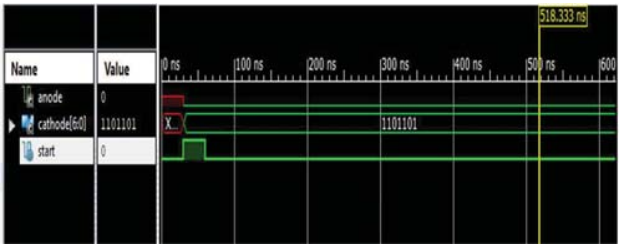


Fig. 12. Digit '5' displayed on Seven-Segment.

```
Time resolution is 1 ps
Simulator is doing circuit initialization process.
Finished circuit initialization process.
Digit predicted is (binary form) :
0101
Anode value of Seven Segment :
0
Cathode value of Seven Segment :
1101101
ISim>
```

Console | Compilation Log | Breakpoints

Fig. 13. SystemVerilog output of the predicted digit '5'.

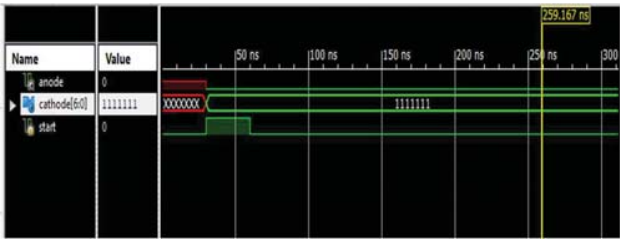


Fig. 14. Digit '8' displayed on Seven-Segment.

```
Time resolution is 1 ps
Simulator is doing circuit initialization process.
Finished circuit initialization process.
Digit predicted is (binary form) :
1000
Anode value of Seven Segment :
0
Cathode value of Seven Segment :
1111111
ISim>
```

Console | Compilation Log | Breakpoints

Fig. 15. SystemVerilog output of the predicted digit '8'.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

```
--------Confusion Matrix--------
Predicted Digit Vs Actual Digit
   0, 1, 2, 3, 4, 5, 6, 7, 8, 9
0 [971, 1, 1, 0, 0, 0, 3, 1, 3, 0]
1 [0, 1127, 2, 2, 0, 1, 1, 0, 2, 0]
2 [3, 2, 1009, 3, 2, 0, 2, 6, 5, 0]
3 [0, 0, 5, 997, 0, 1, 0, 2, 3, 2]
4 [0, 0, 3, 1, 961, 0, 4, 2, 0, 11]
5 [4, 0, 0, 8, 2, 870, 1, 1, 4, 2]
6 [2, 2, 1, 1, 4, 3, 942, 1, 2, 0]
7 [2, 9, 6, 8, 1, 0, 0, 994, 0, 8]
8 [3, 1, 2, 7, 1, 3, 2, 2, 949, 4]
9 [2, 2, 0, 7, 7, 4, 2, 2, 2, 981]
```

Fig. 16. Confusion Matrix (10000 Testing Digits).

X. CONCLUSION

The paper has proposed an algorithm for fingertip writing digit recognition system using CNN. Proposed recognition system processes both pre-processed image algorithms and fingertip writing numbers recognition algorithms in hardcode. The CNN model has been trained on 50,000 handwritten digits provided by an MNIST and has been tested on 10,000 images. The computational efficiency of this algorithm is very high as it implements FPGA for digit recognition and the accuracy of digit recognition is 98.4%. The proposed algorithm makes use of both SystemVerilog as well as python for fingertip hand recognition. SystemVerilog is used just to communicate between hardware and software. Direct programming interface facility of SystemVerilog has been used to read an image and to convert it in its RGB value using C program. That corresponding RGB value has been further passed on to Python program where digit recognition took place and the recognised digit was passed back to SystemVerilog via C program. The future expectation of this algorithm is the betterment of the recognition system and increasing its computational efficiency manifold. In our coming papers, we will be proposing the algorithm to implement the CNN on hardware only as it will increase the computational efficiency and will make the algorithm faster with a high accuracy rate. So that our main motive of helping humankind by making use of machines will come true.

ACKNOWLEDGMENT

We express our deepest gratitude to all those who helped us make this research complete and meaningful in all sense. First of all, we would like to thank our faculty in-charge, Prof. S. Indu, whose contribution to this research work is incredible. She helped us to get through this research work by providing her valuable suggestions and sparing her precious time. Furthermore, we acknowledge the contribution of other staffs of the ECE department (DTU), who helped us by providing their valuable suggestions related to our topic. Last but not least, we are highly thankful to our parent and friends, who provided us moral support in times of distress and failure and played a vital role to make this work complete in all sense.

REFERENCES

[1] Ching-Long Shih, et al, (2015), FPGA Implementation of Vision-Based Fingertip-Writing Digits Recognition System, IEEE International Conference on Systems, Man, and Cybernetics.

[2] O.D. Trier, et. Al, Feature extraction methods for character recognition – a survey, Pattern Recognition, Vol. 29, No 4, pp. 641-662, 1996.

[3] R. Plamondon, S. N. Srihari, On-line & off-line Handwriting Recognition: A Comprehensive Survey, IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol 22, No 1, pp. 63-84, 2000.

[4] Sohom Mukherjee, et al, (2018), Fingertip Detection and Tracking for Recognition of Air-Writing in Videos.

[5] K. Outcharov, et. Al, Accelerating Deep Convolutional Neural Networks Using Specialized Hardware, (Feb 2015)

[6] V. R. Rudra., K. Bhavanish., Handwritten Digit Recognition using CNN, Int. J. of Innovative Sc. & Research Tech., Vol. 4, Iss. 6.

[7] L.S.Oliveira, R. Sabourin, Support vector machines for handwritten numerical string recognition, Int.Workshop on Frontiers in Handwriting Recognition, Tokyo, Japan (2004).

[8] Iuliana Chiuchisan, (2013), A New FPGA-based Real-Time Configurable System for Medical Image Processing, 4th IEEE International Conference, Romania.

[9] Ahinus H, et al, A New Object Detection and Tracking Using FPGA, IOSR Journal of Electronics & Communication Engineering.

[10] Feldman J., Singh M., Information along contours and object boundaries, Psychological Review, Vol. 112, pp. 243-252.

[11] Alan L., Tomas V., Luka C., Jiri M., Matej K. (2016). Discriminative Correlation Filter with Channel and Spatial Reliability. Int. J. of Computer Vision, 126(8).

[12] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, F. Porikli, L. Cehovi, G., G. Fernandez and T. Vojir. 2013 The Visual Object Tracking, The Visual Object Tracking Challenge Results, Page 98-111.

[13] X. Yang, J. Pu, MDig: Multi-digit Recognition using Convolutional Neural Network on Mobile, 14th International Conference on Natural Computation, Huangshan, China.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:05:10 UTC from IEEE Xplore. Restrictions apply.

Handling Imbalanced Data using Ensemble Learning in Software Defect Prediction

Ruchika Malhotra
Computer Science and Engineering Department
Delhi Technological University
Delhi, India
Ruchikamalhotra2004@yahoo.com

Juhi Jain
Computer Science and Engineering Department
Delhi Technological University
Delhi, India
erjuhijain@gmail.com

***Abstract*—With the ever growing software industry, software defect prediction is one of the key ingredients in recipe of producing good quality software. Defects uncovered well in time helps in saving resources in terms of time, effort and money. However imbalanced nature of software data may hamper the resultant performance of models leading to incorrect interpretations of results. This problem has dragged attention of researchers and many solutions are proposed to overcome the effect of this problem. This paper aims to provide empirical comparison of software defect prediction models developed by using various boosting based ensemble methods on three open source JAVA projects. Four ensemble methods incorporate resampling techniques within them. Performances of models obtained are evaluated using stable metrics like Balance, G-Mean and AUC. Results show that use of resampling techniques before classifying using ensemble method has significantly improved model prediction as compared to classic boosting models. RUSBoost is the undisputed winner amongst all followed by MSMOTEBoost and SMOTEBoost.**

Keywords—software defect prediction; data imbalance problem; resampling techniques; ensemble methods; boosting

I. INTRODUCTION

Software is the heart of this technological era. If there are frequent defects in software, it reduces reliability and credibility of software resulting into unsatisfied customers deteriorating the company's image. If defects are uncovered in early stages of software development, then one can allocate resources to required software component only. So, early detection of defects not only save time and resources but also enhances company's goodwill increasing its market share [1]. Software defect prediction using machine learning has therefore gained researchers' attention in recent years with aim of timely delivery of softwares with optimal utilization of resources. But one of the major concerns in machine learning is dealing with imbalanced nature of software data [2]. Considering binary problems, accuracy is considered a good performance metric for comparison of models but what if number of instances of one class is much greater than that of another class? Let us say, out of 100 instances 90 are defect free and 10 have defects. When model is trained on such imbalanced data, will it affect the validation results? Of course, the results will be biased towards majority class. This class imbalance problem is prevalent in many real world problems. One possible solution to this problem is use of ensemble methods. Aim of this study is to provide empirical comparison of predictive capabilities of machine learning models that are developed using broadly two types of ensemble methods taking C4.5 as base classifier. Research questions designed to achieve the addressed objective are:

RQ1: Do resampling techniques enhance the performance of boosting based ensemble methods in case of imbalance data in software world?

RQ2: Which boosting based ensemble method outperforms for prediction of software defects?

Static code attributes of three open source JAVA softwares are exploited to predict software defects. Three classic boosting ensembles used in this study are AdaBoost, AdaBoost.NC and AdaBoost.M1. Performances are then measured in terms of stable metrics- Balance, G-Mean and Area under Curve (AUC) and compared with that of four boosting based ensembles that involve data balancing strategies (RUSBoost, SMOTEBoost, MSMOTEBoost, DATABoost). Estimates of performance evaluators are gained by performing ten-fold cross validation for each dataset. Static code attributes of three open source JAVA softwares are exploited to predict software defects. Statistical validation of models developed is executed using nonparametric test-Friedman Test.

Rest of paper is arranged in four more sections. Section II reviews the related studies of this work. Research methodology and experimental design is examined in section III. Section IV discusses and analyses the results. Last, but not the least, section V concludes the empirical study performed and reveals the future directions.

II. RELATED WORK

Class imbalance problem is encountered in large number of real world problems. Some of them are fraud detection [3], risk management [4] and fault diagnosis [5] [6]. Object oriented (OO) metrics are widely used to explore software quality attributes in existing literature [7]. This work will be focused on application of boosting based ensembles using these static code attributes. Many algorithms are designed to handle imbalanced data. Tosun et al. [8] presented ensemble method that outperformed naïve bayes using balance as performance metric for NASA datasets. Sun et al. [9] proposed new

978-1-7281-2791-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:49:44 UTC from IEEE Xplore. Restrictions apply.

technique for handling class imbalance by converting binary class problem into multiclass problem. Wang and Yao [10] also analyzed different solution to class imbalance problem including resampling and ensembles over the NASA datasets and proposed a new variant of AdaBoost.NC. A novel overbagging based framework has been proposed for NASA datasets by Mousavi et al. [11] in which ensemble is dynamically selected. Huda et al. [12] also proposed the ensemble method based on three oversampling methods and performed comparative analysis with some existing strategies based on accuracy, recall and AUC. Song et al. [13] conducted the detailed experimentation and have included open source software data but not explored the performance based on Balance and AUC. Closest study to our work is Rodreguiz et al. [14] but they performed empirical comparison on NASA datasets where as we have used open source software projects. Also, they did not explore G-Mean and Balance performance evaluators that are considered stable in imbalanced data world.

III. RESEARCH METHODOLOGY

In this section, we notify about the source and nature of our datasets and dependent/independent variables of study followed by the incorporated data preprocessing steps. Next, various ensemble methods used to generate prediction models, performance evaluators employed to compare those models and statistical tests exploited to validate the models are explained.

A. Dataset Collection

This study exploits three open source Apache projects that are coded in JAVA language. Datasets used are curated by Marian Jureckjo [15] and are available on promise repository. Three datasets selected are Synapse 1.0, Camel 1.4 and Ant 1.0. Apache Synapse 1.0 is service oriented architecture (SOA) framework that provides web services. Camel 1.4 is JAVA framework that assists users to implement various design patterns for enterprise integration where as Ant 1.7 is flexible JAVA library used to build JAVA and non JAVA applications. Table I summarizes the datasets in terms of class distribution in increasing order of percentage of defective classes. Total number of classes in software is denoted by #Classes. #DC corresponds to number of defective classes and #NDC stand for number of defective classes. %age of DC signifies the actual proportion of defective classes in complete software. As is evident from table I, %age of DC varies from 10.19% to 22.28%. This represents highly imbalanced nature of software data.

B. Independent and Dependent Variables

Static code attributes viz. OO metrics of the project describes the internal software quality attributes and are used as independent variables in this study. In above described datasets, there are 21 independent variables comprising of CK metric suite [16], five metrics from QMOOD metric suite [17], two metrics from Martin [18], one metric from Henderson-Seller [19], three metrics based on McCabe's complexity, LOC metric and three additional metrics. Metrics' definitions can be referred from http://gromit.iiar.pwr.wroc.pl/p_inf/ckjm/metric.html.

TABLE I. DATASET DETAILS

Dataset	#Classes	#DC	#NDC	%age of DC
Synapse1.0	157	16	141	10.19
Camel1.4	872	727	145	16.63
Ant1.7	745	166	580	22.28

Defect is the binary dependent variable identified for this experimentation. Defect variable can have two possible values- 0 for nondefective class and 1 for defective class. Original data set includes number of defects which can have multiple values. We convert it into binary variable-defect as desired for the purpose of our study.

C. Data Preprocessing

Datasets were checked for any missing data. Then two step data preprocessing is carried out for the datasets. First, outliers are identified and removed. Second, important features are selected amongst 21 OO metrics. There was no missing data in related datasets.

- Outlier Removal: If value of any instance lies far away from range of values of other instances, then that instance is considered outlier. Outliers are located in this study with help of inter quartile range. We used Waikato Environment for Knowledge Analysis (WEKA) tool for this purpose [20]. We identified 53, 50 and 125 outliers in Synapse 1.0, Camel 1.4 and Ant 1.7 respectively. We removed these outliers so that model validation is not affected by extremities.
- Feature Selection: After the outliers are removed, we focused on selecting the important features of the datasets. Performance of classifier is hampered by the presence of irrelevant or redundant features in feature space [21] [22]. We used correlation feature selection (CFS) that is proposed by [23]. Reference [22] also favors the application for CFS for measuring the goodness of features in imbalanced data. Selected features for respective datasets are noted in table II.

D. Ensemble Methods

Though there are many solutions to class imbalance problem, this study is limited to boosting algorithms in ensemble methods. Many classifiers are trained using training data and concluding judgment is based on their combination. C4.5 [25] is taken as base classifier for generating ensemble models. Boosting algorithm aims to improve the predictive capability of learner by recognizing the weak spots, i.e., instances that are wrongly classified. It amplifies the weightage of incorrect classified instances and new learner utilizes this improved dataset. Boosting creates several models that harmonize each other and prediction is done based on their weighted voting. AdaBoost is adaptive boosting algorithm that works on sampling framework [27] and its variant AdaBoost.NC is also incorporated. Adaboost.NC is adaptive boosting with negative correlation learning extension.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:49:44 UTC from IEEE Xplore. Restrictions apply.

AdaBoost.M1 is another extension of AdaBoost for binary class problem.

TABLE II. SELECTED FEATURES OF DATASETS USING CFS

Dataset	Features Selected
Synapse 1.0	LOC, MAX_CC, AMC, RFC, WMC
Camel 1.4	WMC, CBO, RFC, Ca, Ce, NPM, IC,CBM
Ant 1.7	WMC, RFC, LOC, CAM, AMC

In AdaBoost.M1 (AdaBM1) [26] ensemble method, as training starts, equal weights are allocated to training data. Weight of the instance is increased or decreased in next iteration depending on whether that instance is correctly classified or not. AdaBoost.M2 (AdaBM2) [27] is a slight variant of previous one which removes the restriction of maximum weight assigned to the instance. RUSBoost (RUSB) [28] is hybrid approach that employs random under sampling of instances before applying boosting. Other preprocessing techniques used are SMOTE, MSMOTE and data resampling. These corresponds to SMOTEBoost (SMOTEB) [29], MSMOTEBoost (MSMOTEB) [30] and DataBoost (DataB) [31] ensemble methods. SMOTEBoost is boosting algorithm based on oversampling technique- Synthetic Minority Oversampling Technique (SMOTE). Synthetic instances of minority classes are created and injected at each boosting iteration. RUSBoost works same as SMOTEBoost. Only difference is in resampling technique used. At each iteration of boosting, instead of SMOTE, random undersampling (RUS) is used. In DataBoost ensemble method, data generation is combined with boosting algorithm to improve the predictive capability of majority as well as minority class. First, hard instances are identified and then synthetic samples are generated based on these hard instances. These synthetic instances are reweighted and then added to training set of data.

E. Performance Evaluators

Many researchers have advocated that accuracy as a performance evaluator do not reflect the true classification picture for imbalanced data [32] [33]. So we have used more stable metrics like G-Mean, Balance and AUC. We also calculated one of traditional metric- sensitivity as it assists in understanding imbalanced nature of data. These metrics are calculated using confusion matrix comprising of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

- Sensitivity: It is percentage of correctly predicted defective classes and is calculated as $Sensitivity = \frac{TP}{TP+FN}$ (1)
- G-Mean: is geometric mean of sensitivity and specificity and is calculated as $GMean = \sqrt{Sensitivity\ X\ Specificity}$ (2)
- Balance: In ideal case, precision should be 1 and false positive rate (FPR) should be 0. So, Balance uses this definition to make a reliable performance evaluator for imbalanced data. It is calculated as

$$Balance = 1 - \sqrt{\frac{(0 - FPR)^2 + (1 - Precision)^2}{2}} \quad where$$
$$FPR = \frac{FP}{FP + TN} \; and \; Precision = \frac{TP}{TP + FN} \qquad (3)$$

- AUC: Area under Receiver operating characteristic (ROC) curve defines relationship between sensitivity and specificity and is well accepted performance metric.

F. Validation and Statistical Tests

Ten fold cross validation is used to evaluate each developed model. In this, a particular dataset is divided into ten approximately equal parts. Nine parts are used for training the model and remaining part is used for testing the model. This procedure is repeated for all ten partitions [34]. Employing ten fold cross validation while training the model builds the confidence in the prediction results as it reduces chance of overfitting.

We also used Friedman test [33] to statistically validate results of models for sensitivity, G-Mean, Balance and AUC. Friedman test is non parametric technique and, therefore, there is no need of any presumption to be satisfied by the datasets. We designed null hypothesis as: there is no difference between the predictive capabilities of seven boosting ensemble methods.

IV. RESULTS AND ANALYSIS

We have used open source Knowledge Extraction based on Evolutionary Learning (KEEL) tool [34] to build required ensemble models for defect prediction available on http://www.keel.es/. During experimentation, default settings are followed to support replication of study. This reduces external validity threats in the study. We executed ten runs for each ensemble method and took average of considered performance evaluators.

A. RQ1: Do resampling techniques enhance the performance of boosting based ensemble methods in case of imbalance data in software world?

All the models were built using KEEL software. Results of employed boosting algorithms are summarized in table III. Performance of models are compared using sensitivity, G-Mean, Balance and AUC. By keen observation, we can see that for sensitivity, range of classic ensembles is from 0.086 to 0.5 whereas boosting algorithm that has resampling technique embedded in it has achieved maximum of 0.75. Similar comparisons can be made for other performance metrics. G-Mean values for classic ensemble methods range from 0.2872 to 0.5461 except for one value. For synapse1.0, AdaBoost.NC has G-mean value 0.6844. Comparatively, G-Mean values for resampling based boosting ensembles range from 0.4024 to 0.7327. 50% of G-Mean values for resampling+boosting ensembles are more than 60%. Maximum Balance value for resampling+boosting ensembles is 0.6709 exhibited by Ant 1.7, which is approximately 20% greater than that achieved by classing boosting ensemble methods. In remaining two datasets also, maximum Balance is shown by RUSBoost. Synapse 1.0

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:49:44 UTC from IEEE Xplore. Restrictions apply.

TABLE III. BOOSTING ENSEMBLES RESULTS

Sensitivity	**Ant1.7**	**Camel1.4**	**Synapse1.0**
AdaB	0.2404	0.1505	0.3333
AdaBNC	0.2596	0.086	0.5
AdaBM1	0.2404	0.1398	0.3333
DataB	0.4808	0.1828	0.4167
MSMOTEB	0.5769	0.3763	0.5
RUSB	0.5769	0.5699	0.75
SMOTEB	0.5288	0.3763	0.4167
G-Mean	**Ant1.7**	**Camel1.4**	**Synapse1.0**
AdaB	0.4674	0.3708	0.5461
AdaBNC	0.4945	0.2872	0.6844
AdaBM1	0.4674	0.3583	0.5461
DataB	0.6418	0.4024	0.6177
MSMOTEB	0.682	0.5562	0.6767
RUSB	0.682	0.6271	0.7327
SMOTEB	0.6646	0.5601	0.5924
Balance	**Ant1.7**	**Camel1.4**	**Synapse1.0**
AdaB	0.459	0.3962	0.5227
AdaBNC	0.4748	0.3531	0.6436
AdaBM1	0.459	0.389	0.5227
DataB	0.6191	0.4165	0.5833
MSMOTEB	0.6709	0.5414	0.6415
RUSB	0.6709	0.6251	0.7324
SMOTEB	0.647	0.5436	0.5727
AUC	**Ant1.7**	**Camel1.4**	**Synapse1.0**
AdaB	0.5758	0.5312	0.6216
AdaBNC	0.6018	0.5223	0.7188
AdaBM1	0.5758	0.5285	0.6216
DataB	0.6696	0.5351	0.6833
MSMOTEB	0.6925	0.5988	0.7322
RUSB	0.6925	0.6272	0.7327
SMOTEB	0.6819	0.6040	0.6472

attained remarkable Balance value of 0.7327 when RUSBoost is applied to it. For AUC, RUSBoost and MSMOTEBoost have given best results for all three datasets. We can conclude that resampling techniques have improved the predictive performance of models in term of G-Mean, Balance, AUC and sensitivity.

B. *RQ2: Which boosting based ensemble method outperforms for prediction of software defects?*

To answer this question, we performed Friedman test to have fair comparison between classic boosting based ensembles and boosting ensembles that have resampling techniques embedded in them. Friedman rankings are highlighted in table IV for seven boosting ensemble methods considering the three selected datasets. It includes ranking allotted to ensembles based on their performance evaluation using sensitivity, G-mean, Balance and AUC. Mean rank attained by particular boosting ensemble method is shown in brackets. Higher the mean rank, better is the predictive capability of the ensemble in term of the concerned performance evaluator. Friedman test is conducted at $\alpha = 0.05$ (95% confidence) and p-values are observed for sensitivity, G-Mean, Balance and AUC. As apparent from table IV, we concluded on following observations:

- P-value for sensitivity and AUC is 0.016 which is less than 0.05. So, we refute our null hypothesis. This shows that there is significant difference between the predictive capabilities of different ensembles using sensitivity and AUC.

TABLE IV. FRIEDMAN RANKINGS FOR ENSEMBLES

Rank	**Sensitivity**	**G-Mean**	**Balance**	**AUC**
Rank 1	RUSB (7.83)	RUSB (7.83)	RUSB (7.83)	RUSB (7.83)
Rank 2	MSmoteB (6.50)	MSmoteB (6.50)	MSmoteB (6.83)	MSmoteB (6.83)
Rank 3	SmoteB (5.67)	SmoteB (5.67)	SmoteB (5.67)	SmoteB (5.67)
Rank 4	DataB (5.00)	DataB (5.00)	DataB (5.00)	DataB (4.83)
Rank 5	AdaBNC (4.00)	AdaBNC (4.00)	AdaBNC (3.67)	AdaBNC (3.83)
Rank 6	AdaB (2.67)	AdaB (2.67)	AdaB (2.67)	AdaB (2.67)
Rank 7	AdaBM1 (2.17)	AdaBM1 (2.17)	AdaBM1 (2.17)	AdaBM1 (2.17)
p-Value	0.016	0.023	0.023	0.016

- Similarly, p-value for Balance and G-Mean is less than 0.05 which indicates significant difference between the model performances.
- First four rankers are the boosting based ensemble methods that use resampling technique before building the model. This strengthens our answer to RQ1. Use of resampling techniques before building ensemble definitely enhances the predictive capability of our software defect prediction models.
- RUSBoost has outperformed for all the datasets when we assessed their predictions using G-Mean, Balance, AUC and sensitivity, followed by MSMOTEBoost, SMOTEBoost and DataBoost.

V. CONCLUSION AND FUTURE WORK

This experimental study provides the empirical comparison of seven boosting based ensembles. Analysis of results advocates the usage of resampling techniques prior to application of boosting algorithm. Their predictive capabilities outperformed the performance of classic boosting algorithms for software defect prediction. All three Apache projects had highly imbalanced ratio of defective and non defective classes, still ensemble methods provide acceptable and significant validation results. We should use G-Mean, Balance and AUC metrics for building models when dealing with imbalanced data. Resampling techniques must be applied before executing boosting. RUSBoost followed by MSMOTEBoost and SMOTEBoost ensemble methods resulted in better classification of software defects than other classifiers.

Other than class imbalance problem, there also exists other reasons like small disjuncts, cross validation data distribution and dataset shift that influence the classifier performance [37]. In future, we can address these problems. One can also empirically compare other categories of ensembles like bagging and cost sensitive ensembles with more datasets. This will aid in generalizing the results.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:49:44 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] Malhotra, R., Empirical research in software engineering: concepts, analysis, and applications. Chapman and Hall/CRC, 2016.

[2] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas. "Handling imbalanced datasets: A review." GESTS International Transactions on Computer Science and Engineering 30, no. 1, pp. 25-36, 2006.

[3] Y.-M. Huang, C.-M. Hung, and H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem," Nonlinear Anal. R. World Appl., vol. 7, no. 4, pp. 720–747, 2006.

[4] D. Cieslak, N. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," in IEEE Int. Conf. Granular Comput., pp. 732– 737, 2006.

[5] Z. Yang, W. Tang, A. Shintemirov, and Q. Wu, "Association rule miningbased dissolved gas analysis for fault diagnosis of power transformers," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 39, no. 6, pp. 597–610, 2009.

[6] Z.-B. Zhu and Z.-H. Song, "Fault diagnosis based on imbalance modified kernel fisher discriminant analysis," Chem. Eng. Res. Des., vol. 88, no. 8, pp. 936–951, 2010.

[7] Y. Singh, A. Kaur, and R. Malhotra, "Empirical validation of object-oriented metrics for predicting fault proneness models," Software quality journal, vol. 18, no. 1 , p. 3, 2010.

[8] A. Tosun, B. Turhan, and A. Bener, "Ensemble of software defect predictors: a case study," In Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement, ACM, pp. 318-320, 2008.

[9] Z. Sun, Q. Song, and X. Zhu, "Using coding-based ensemble learning to improve software defect prediction," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) , vol.42, no. 6, pp. 1806-1817, 2012.

[10] S.Wang, and X. Yao, "Using class imbalance learning for software defect prediction," IEEE Transactions on Reliability, vol. 62, no. 2, pp. 434-443, 2013.

[11] Mousavi, Reza, Mahdi Eftekhari, and Farhad Rahdari, "Omni-Ensemble Learning (OEL): Utilizing Over-Bagging, Static and Dynamic Ensemble Selection Approaches for Software Defect Prediction," International Journal on Artificial Intelligence Tools, vol. 27, no. 6, 1850024, 2018.

[12] Huda, Shamsul, Kevin Liu, Mohamed Abdelrazek, Amani Ibrahim, Sultan Alyahya, Hmood Al-Dossari, and Shafiq Ahmad, "An ensemble oversampling model for class imbalance problem in software defect prediction.," IEEE access 6, pp.24184-24195, 2018.

[13] Q. Song, G. Yuchen, and M. Shepperd, "A comprehensive investigation of the role of imbalanced learning for software defect prediction," IEEE Transactions on Software Engineering, 2018.

[14] D. Rodriguez, I. Herraiz, R. Harrison, J. Dolado, and J. C. Riquelme, "Preliminary comparison of techniques for dealing with imbalance in software defect prediction," In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, p. 43, 2014.

[15] T. Menzies, R. Krishna, and D. Pryor D, The promise repository of empirical software engineering data. North Carolina State University, Department of Computer Science [Online]. http://openscience.us/repo

[16] S. R.Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," IEEE Trans Software Eng, vol. 20, no. 6, pp.476-493, 1994.

[17] J. Bansiya and C. G. Davis, "A hierarchical model for object-oriented design quality assessment," IEEE Trans Software Eng, vol. 28, no. 1, pp. 4–17, 2002.

[18] R. C. Martin, "Object oriented design quality metrics: an analysis of dependencies," Rep Object Anal Design, vol. 2, no. 3, pp. 1–8, 1995.

[19] B. Henderson-Sellers, Object-oriented metrics: measures of complexity. Prentice-Hall, Inc., 1995.

[20] F. Eibe, M. A. Hall, and I. H. Witten, The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[21] C. Catal and B. Diri, "Investigating the effect of dataset size, metrics sets, and feature selection techniques on software fault prediction problem," Inf. Sci., vol. 179, no. 8, pp. 1040–1058, 2009.

[22] L. Yin, Y. Ge, K. Xiao, X. Wang, and X. Quan, "Feature selection for high-dimensional imbalanced data", Neurocomputing, vol. 105, pp. 3–11, 2013.

[23] M. A. Hall and L. A. Smith, "Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper," In FLAIRS conference, vol. 1999, pp. 235-239, 1999.

[24] M. Wasikowski and X. W. Chen, "Combating the small sample class imbalance problem using feature selection," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1388-1400, 2009.

[25] J. R. Quinlan, "C4. 5: Programs for machine learning,", Morgan Kaufmann, San Francisco, 1993.

[26] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," In Thirteenth International Conference on Machine Learning, San Francisco, Morgan Kaufmann, pp 148–156, 1996.

[27] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp.119–139, 1997.

[28] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 40, no. 1, pp.185–197, 2010.

[29] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," In European conference on principles of data mining and knowledge discovery, Springer, Berlin, Heidelberg, pp. 107-119, 2003.

[30] S. Hu, Y. Liang, L. Ma, and Y. He, "MSMOTE: improving classification performance when training data is imbalanced," In 2009 second international workshop on computer science and engineering, vol. 2, pp. 13-17. IEEE, 2009.

[31] S. Wang and Y. Xin, "Using class imbalance learning for software defect prediction," IEEE Trans on Reliability, vol. 62, no. 2, pp.434-443,2013.

[32] H. Guo, and H. L. Viktor, "Learning from imbalanced data sets with boosting and data generation: the databoost-im approach," ACM Sigkdd Explorations Newsletter vol. 6, no. 1, pp. 30-39, 2004.

[33] H. He and E. A. Garcia. "Learning from imbalanced data." IEEE Transactions on knowledge and data engineering vol. 21, no. 9, pp. 1263-1284, 2009.

[34] I. Myrtveit, E. Stensrud, and M. Sheppered, "Reliability and validity in comparative studies of software prediction models", IEEE transactions on Software Engineering ,vol. 31, no.5, pp.380-390, 2005.

[35] M. Friedman, Milton, "A comparison of alternative tests of significance for the problem of m rankings," The Annals of Mathematical Statistics 11, no. 1, pp. 86-92, 1940.

[36] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez and F. Herrera, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework," Journal of Multiple-Valued Logic & Soft Computing, vol. 17, 2011.

[37] V. L´opez, A. Fern´andez, and F. Herrera, "On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed," Information Sciences, vol. 257, pp. 1–13, , 2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:49:44 UTC from IEEE Xplore. Restrictions apply.

2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW)

Hierarchical Clustering of World Cuisines

1st Tript Sharma
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
triptsharma22@gmail.com

1st Utkarsh Upadhyay
Department of Electronics Engineering
Jamia Millia Islamia University
New Delhi, India
utkarshdhy@gmail.com

2nd Jushaan Kalra
Department of Computer Engineering
Delhi Technological University
New Delhi, India
jushaan18@gmail.com

3rd Sakshi Arora
Department of Computer Science
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
sakshi18133@iiitd.ac.in

4th Saad Ahmad
Department of Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
saad18409@iiitd.ac.in

4th Bhavay Aggarwal
Department of Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
bhavay18384@iiitd.ac.in

Ganesh Bagler
Center for Computational Biology
Indraprastha Institute of Information and Technology
New Delhi (IIIT-Delhi), India
bagler@iiitd.ac.in

***Abstract*—Cultures across the world have evolved to have unique patterns despite shared ingredients and cooking techniques. Using data obtained from RecipeDB, an online resource for recipes, we extract patterns in 26 world cuisines and further probe for their inter-relatedness. By application of frequent itemset mining and ingredient authenticity we characterize the quintessential patterns in the cuisines and build a hierarchical tree of the world cuisines. This tree provides interesting insights into the evolution of cuisines and their geographical as well as historical relatedness.**

***Index Terms*—Hierarchical Clustering, Pattern Mining, Authenticity Correlation, Kmeans Clustering, Food Ontology**

I. Introduction

Cultures across the world have evolved diverse cooking practices over time. Although the underlying fundamentals of cooking remain the same, various factors including geography and climate have affected cooking styles. Cuisines from across the globe have thus acquired their signature styles. Each cuisine has interesting patterns that are inherent to it while sharing some common attributes with others. In this article, we characterize the unique features that typify every cuisine in an attempt to discern the footprint of food on human cultures and inter-relatedness of world cuisines.

II. Literature Survey

With increasing availability of recipes data, there has been much interest in data mining recipes data. One of the focus has been on defining recipe similarity. Attempts have been made to define similarity based on various elements of cooking recipes [15] and ingredients [7], [14].

Among other efforts in data mining recipes data have focused on food pairing phenomena in cuisines. Among one of the early studies, Shidochi et al [11] experimented with the possible replacements of ingredients in a recipe. Jain et al [8] investigated the phenomenon of food pairing which examines compatibility of two ingredients in a recipe in terms of their shared flavor compounds. This study investigated the food pairing phenomena in Indian recipes and proclaimed that spices form the basis of their food pairing. The work was extended by Singh et al [12] to analyze a much larger dataset encompassing 22 cuisines across the world and found interesting food pairing patterns in cuisines from across the world. Tuwani et al [13], on the other hand, considered culinary systems as a function of socio-cultural factors and presented computational models for cuisine evolution. An interesting work by Yokoi et al [16] calculated an ingredient associative metric called 'typicality value', giving out typical recurring ingredient patterns.

In this article, we indulge in pattern analysis in world-wide cuisines by way of association rule discovery and frequent pattern mining [1]. Going beyond the application of pattern mining techniques on cuisines in [10], we propose their use for frequent pattern mining of recipe data and cooking processes, utensils and ingredients for hierarchical clustering of cuisines.

III. Data Collection

Our analysis involved four types of information pertaining to traditional recipes, namely, recipes, ingredients, processes and utensils. A total of 118,071 recipes were obtained from various sources: AllRecipes, Food Network, Epicurious and TarlaDalal. RecipeDB [3], a structured compilation of recipes, was used as the primary source of information. All data are available at 'RecipeDB: A resource for exploring recipes'.

For each recipe, details such as its name and the list of ingredients and processes involved while cooking are

2473-3490/20/$31.00 ©2020 IEEE
DOI 10.1109/ICDEW49219.2020.00007

TABLE I
SIGNIFICANT PATTERNS MINED FROM CUISINES ACROSS THE WORLD

Region	Number of Recipes	Pattern	Support	Number of patterns
Australian	5,823	Butter	0.24	29
Belgian	1,060	Butter + salt	0.24	51
Canadian	6,700	Onion	0.20	31
Caribbean	3,026	Garlic Clove	0.24	32
Central American	460	Onion	0.30	38
Chinese and Mongolian	5,896	Soy sauce + add + heat	0.27	88
Deutschland	4,323	Onion	0.29	54
Eastern European	2,503	Cream	0.30	60
French	6,381	skillet	0.21	60
Greek	4,185	Olive Oil	0.40	43
Indian Subcontinent	6,464	Onion + add + heat + salt	0.22	119
Irish	2,532	Butter	0.32	41
Italian	16,582	Parmesan cheese	0.31	63
Japanese	2,041	Soy Sauce	0.45	45
Mexican	14,463	cilantro	0.25	33
Rest Africa	2,740	Onion + add + heat	0.20	51
South American	7,176	Onion + salt	0.21	62
Southeast Asian	1,940	Fish sauce	0.24	69
Spanish and Portuguese	2,844	Olive Oil	0.31	67
Thai	2,605	Fish sauce + add + heat	0.23	73
Korean	668	Soy sauce + sesame oil green onion + sesame oil	0.34 0.24	85
Middle Eastern	3,905	Salt + bowl Lemon Juice	0.22 0.22	46
Northern Africa	1,611	cumin + cinnamon cumin + olive oil cumin + Salt	0.21 0.22 0.22	134
Scandinavian	2,811	Butter + Salt Salt + Sugar	0.22 0.21	52
UK	4,401	Butter Oven	0.37 0.46	45
US	5,031	Bake + preheat+ oven + bowl Onion	0.22 0.25	67

available. Each recipe was treated as an unordered list of ingredients, processes and utensils. We integrated recipes from all the sources and grouped them into 26 distinct geo-cultural 'cuisines' while ensuring that each region had enough recipes attributed to it to distinguish it as a cuisine. Please refer to Table I for the list of regions. Due to insufficient information about the region for many recipes, they were aggregated together on the basis of their geological similarities with the prefix 'Rest'. For example, recipes without 'region' information belonging to Africa were put in 'Rest Africa' category.

The database consists of 20,280 unique ingredients, 268 unique processes and 69 unique utensils. The data are sparse in the list of utensils and 14,601 recipes don't have information regarding the preferred utensils required for cooking. An average recipe in a cuisine has $\sim$10 ingredients, $\sim$12 processes and $\sim$3 utensils. This is intuitive as too many ingredients would impede the success/propagation of a recipe, whereas too few would lead to it being modified easily. Thus recipes needs to maintain a balance between complexity and simplicity to survive successive iterations of evolution [13].

IV. PATTERN MINING

To investigate the ontology of food, we mined rules from the data to understand the patterns that are prevalent in a given cuisine. The methodology employed for mining patterns is explained in Section V. The mined patterns consist of ingredients, processes and utensils permutations that have a frequency greater than the defined threshold support. According to [1], [6], support represents the frequency with which the collection of items co-occur as a percentage of all transactions. A high support threshold represents high confidence in the pattern being mined, whereas with a low support threshold noise can creep into the mined patterns, leading to false identification of cuisine features. Hence, a trade off support of 20% was chosen as the threshold. Since the mined patterns are the most frequent ones, it is safe to say that most of the recipes follow the observed patterns and essentially define the cooking practices of the cuisine.

Among the patterns obtained in all the recipes for the Korean region as shown in Table I, the pattern "Soy Sauce + sesame oil" occurs with a support of 0.34, i.e. the pattern is found in 34% of all the recipes in the Korean region. Table I contains the topmost significant patterns in the 26 cuisines. The pattern depicts set of words occurring in a particular

recipe. The patterns mined are highly skewed, with most regions containing patterns having generic ingredients such as 'salt', 'onion' and processes such as 'add' and 'cook', which is justified as they have a high frequency among all cuisines and are fundamental to cooking in many cuisines.

V. Methodology

We implemented two approaches namely Frequent Itemset-based Hierarchical Clustering (FIHC) [5] and Authenticity based Clustering [2] to extract relationships between various cuisines of the world. The hypothesis is that some patterns which are common across a subset of cuisines would be found, which defines their 'similarity/closeness' with each other.

A. Frequent Pattern Mining

Frequent Itemset Mining refers to discovering interesting patterns in databases such as association rules from a set. Since we treat a recipe as a combination of ingredients, processes and utensils, it can be treated as an unordered set of these entities. For the frequent itemset mining, the FP-Growth Algorithm [6] was used as it is an efficient and scalable method for mining the complete set of frequent patterns by pattern fragment growth. The data extracted from RecipeDB was pre-processed to make it compatible with the input form of FP-Growth Algorithm. Ingredients, utensils and processes were concatenated and the FP-Growth Algorithm was applied. This approach was applied to all 26 regions present in the data extracted from RecipeDB. The support was kept at 0.2 so that the pattern was mined across a reasonable number of recipes.

B. Authenticity Based Clustering

We propose that a cuisine can be represented as a set of ingredients, process and utensils which can thus be utilized to define the relationships among the cuisines. Using the authenticity metric described in [2] we calculate the prevalence P_i^c of an item i in a cuisine c according to equations 1 as a function of number of recipes, n_i^c in a cuisine over total number of recipes in the dataset, N_C. This is used to calculate the authenticity of the item for a cuisine using equation 2.

$$P_i^c = n_i^c / N_C \tag{1}$$

$$p_i^c = P_i^c - (P_i^k)_{c \neq k} \tag{2}$$

In order to obtain the contribution of the item in uniquely identifying a cuisine, a relative prevalence matrix is created by subtracting the average prevalence of the item, say i for all cuisines from the prevalence for cuisine c. Accordingly, the most prevalent and least prevalent items in a cuisine can be identified. It should be noted that both the most prevalent and least prevalent items would contribute towards the culinary fingerprint of a cuisine as the former indicates the items having a relatively higher utility in the cuisine while the latter indicates items that are least used in the cuisine versus the rest of the world cuisines.

VI. Clustering Techniques

A. Hierarchical Clustering

Application of FP-Growth Algorithm on the prepared dataset results in 26 files, each containing patterns in a 'frozenset' along with their respective support to remove redundant patterns. These patterns were extracted from the 'frozenset' and appended together in a list in a sorted fashion. All the elements of this list are appended and converted into a string resulting in a 'string pattern'. All the 'string patterns' are compiled into a set resulting in unique set of patterns across all the 26 regions. Since the data is in string form and each element is a unique entity, it can be classified as a category. Therefore, unique set of 'string patterns' are fit for using Label Encoding (because the strings are categorical data) to get a transformer and the 'string patterns' in the rules are transformed using the derived transformer across all the regions. All the 'string patterns' in the rules from all the regions are appended in an array. The data from this array is thus vectorized to form a feature vector which is thus fed to the cluster as the linkage matrix.

Three different approaches were applied in order to cluster the linkage matrix data and to generate subsequent dendrograms. The linkage matrix is converted into a condensed distance matrix (pdist) in order to calculate the distance between all the cuisines based on the rules mined and is then fed into the hierarchical clustering model. To analyze the clusters we have used three distance metrics:

$$\text{Jaccard Distance} = \frac{c_i \cup c_j}{c_i \cap c_j} \tag{3}$$

$$\text{Cosine Distance} = \frac{c_i . c_j}{|c_i||c_j|} \tag{4}$$

$$\text{Euclidean Distance} = \sqrt{c_i^2 + c_j^2} \tag{5}$$

where cuisines c_i, $c_j \in C$, the universal set of cuisines in the dataset. To calculate the distance between two cuisines, they must be quantified. This was done by vectorizing the patterns obtained by the above-mentioned pre-processing technique.

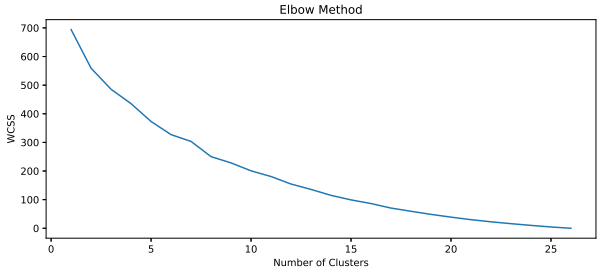


Fig. 1. Elbow Method for cluster identification

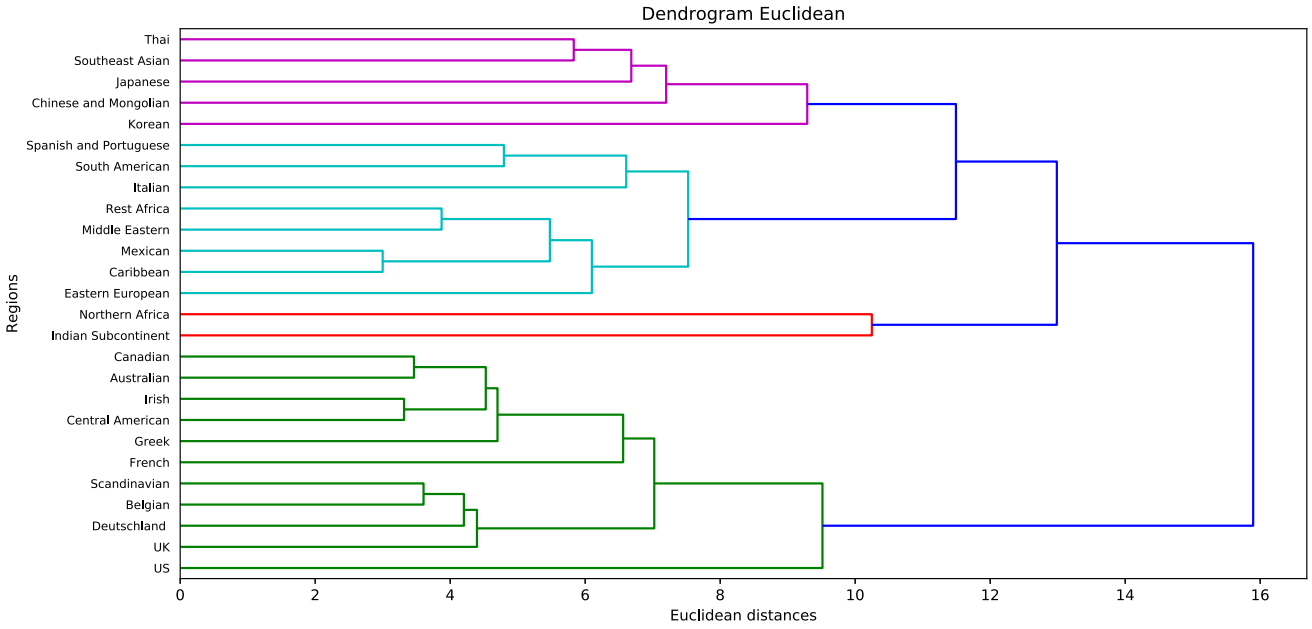


Fig. 2. Hierarchical Agglomerative Clustering based on Patterns Mined using Euclidean distance

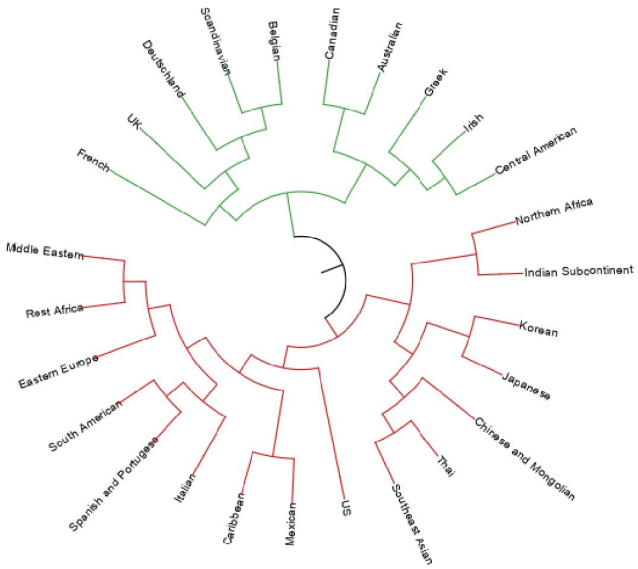


Fig. 3. Hierarchical Agglomerative Clustering based on Patterns Mined using Cosine distance

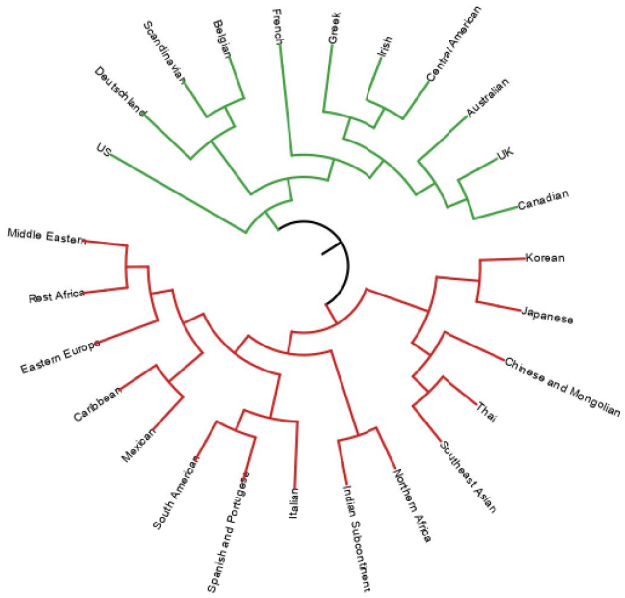


Fig. 4. Hierarchical Agglomerative Clustering based on Patterns Mined using Jaccard distance

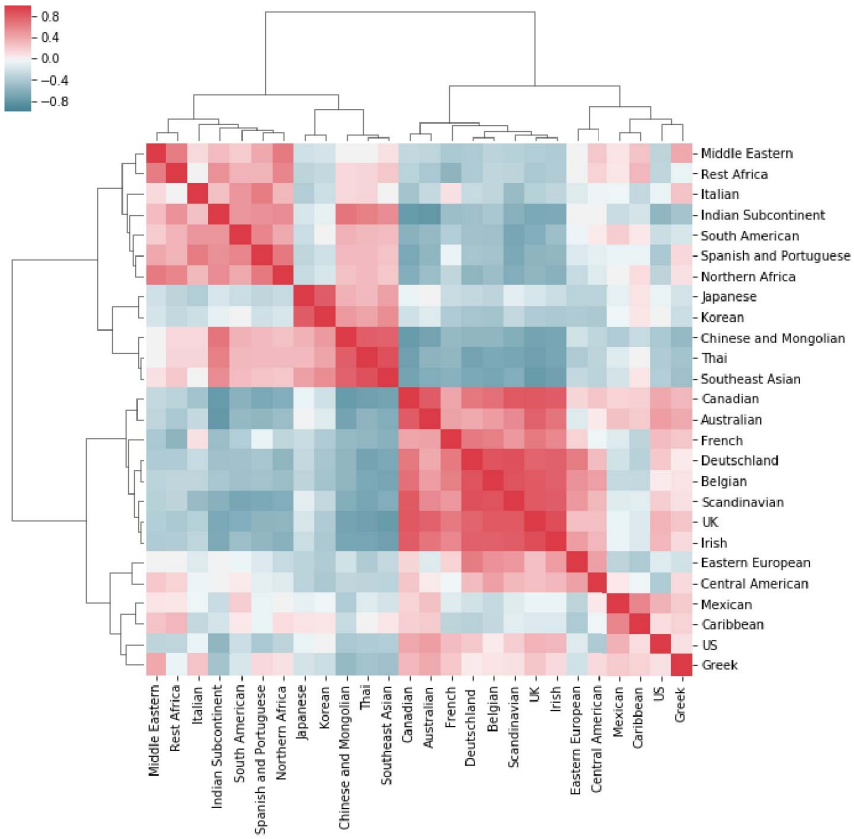


Fig. 5. Hierarchical Agglomerative Clustering based on Authenticity of Ingredients

B. *K-means Clustering*

Another popular clustering technique, K-means, was applied on our categorical data. It has been shown in [9] that hierarchical agglomerative clustering is a better approach for clustering categorical data than K-means. The elbow analysis and the subsequent WCSS score on our dataset indicates similar results. The elbow method [4] analysis fails to determine the number of appropriate clusters for our dataset. As in Figure 1, no sharp edge or elbow like structure is obtained which determines the number of clusters. While on the other hand the hierarchical agglomerative clustering technique presents with better cluster representation. Therefore, our results were predominantly determined by hierarchical agglomerative clustering.

VII. Results

To evaluate the efficiency of the proposed methodologies, the RecipeDB dataset mentioned in Section III was used. This dataset was used to identify the patterns which were then fed into the Sequential Pattern Mining based clustering algorithm while the ingredients obtained from the dataset were the input features for the Authenticity-based clustering. The corresponding code and relevant files are present in the GitHub repository (https://github.com/cosylabiiit/Hierarchical-Clustering-Ingredients).

The Hierarchical Agglomerative clustering (HAC) gives clusters for all regions based on the three approaches and presents a cluster dendrogram for each approach. Figures 2, 3 and 4 represent the clusters formed using the feature vector obtained via Euclidean, Cosine and Jaccard metrics for pairwise distance calculation respectively. Similarly, Figure 5 shows the authenticity based approach to determine the correlation of cuisine and regions, dominantly based on ingredients.

Because of the absence of a quantified validation metric for cuisine similarity, the geographical relationship among the cuisines was used to validate the accuracy in the prediction of cuisine interrelationships. It is observed that while comparing the Figures 2, 3 and 4 with Figure 6 the results received from the Euclidean distance model were most similar to the geographical distribution of the countries. On the other hand, the clusters obtained via the authenticity based clustering gave similar yet better results than Euclidean distance-based HAC when validated on geographical distance based clusters.

Authenticity-based Clustering identifies both positive and negative relationships between cuisines and items whereas pattern based techniques take only the positive relationships

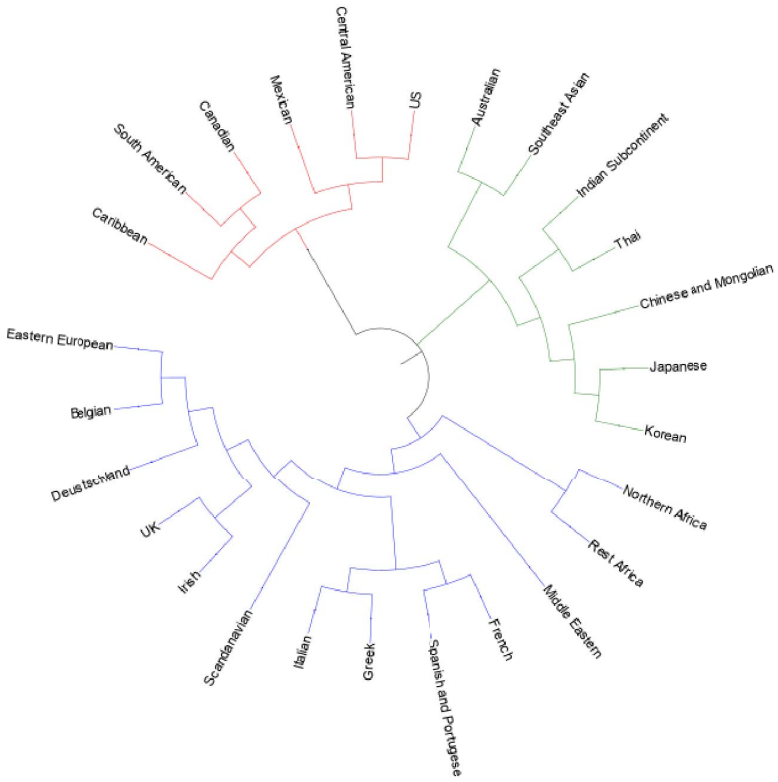


Fig. 6. Hierarchical Agglomerative Clustering based on Geographical Distance of Regions

into account. This leads to the difference in the results. Yet, despite the differences, both techniques predict a closer relationship among Canadian and French cuisines as compared to Canadian and US cuisines despite their geographical proximity. This is evident from the historical fact that Canada was a French colony. Another interesting grouping is that of Indian Subcontinent and Northern Africa. Due to prevalent use of spices in the two regions, Indian subcontinent cuisine is closer to African cuisine as compared to its geographical neighbors like Thai and Southeast Asian cuisines. Hence, the obtained clusters are also able to identify relationships deviating from the geographical similarities.

VIII. Conclusions and Future Work

In this exploratory work we proposed and analyzed two methodologies for fingerprinting cuisines and identifying their interdependence. Our clustering algorithms show how various cuisines are interrelated and show trends similar to their geographical associations. It shows how cooking practices and methods are shared by neighbouring regions. This analysis is important from a historical and cultural point of view as it helps in appreciating how cooking practices are distributed across the world. Furthermore, we also provide a verbose list of patterns identified in the cuisines. These patterns include compound patterns; combination of ingredients, processes and utensils that can be used to identify the relationship among these items.

While this article introduces new methods for investigation of cuisine correlations, it raises new research questions. How do factors such as climate, economy and genetics influence the cuisine patterns? RecipeDB is a sparse dataset in terms of utensils and processes. Hence, to what extent do they influence the relationships among cuisines is yet to be answered. Among one of the limitations of this study, it neither considers the state of ingredients nor their aliases. Therefore, future analysis need to account for the aliases along with state of ingredients and other properties like cooking time and preparation time for the task. It would also be interesting to identify more sophisticated validation metric for cuisine ontology than geographical clustering.

We believe that this study can be applied for cuisine fingerprinting, food ontology and exploration of relations between food and culture. Probing the past and present interrelatedness among cuisines can provide insight into human behavior and

cultures, and means for shaping the future of food.

IX. Acknowledgement

G.B. thanks the Indraprastha Institute of Information Technology (IIIT-Delhi) for providing computational facilities and support. T.S, U.U, J.K,S.A, S.A. and B.A. are Research Interns in Dr. Bagler's lab (Complex Systems Laboratory) at the Center for Computational Biology. All the research interns are thankful to IIIT-Delhi for the support.

References

[1] R. Agrawal, R. Srikant *et al.*, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, 1994, pp. 487–499.

[2] Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, and A.-L. Barabási, "Flavor network and the principles of food pairing," *Scientific reports*, vol. 1, p. 196, 2011.

[3] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh *et al.*, "Recipedb: A resource for exploring recipes," *Available at SSRN 3482237*, 2019.

[4] M. J. Brusco and J. D. Cradit, "A variable-selection heuristic for k-means clustering," *Psychometrika*, vol. 66, no. 2, pp. 249–270, 2001.

[5] B. C. Fung, K. Wang, and M. Ester, "Hierarchical document clustering using frequent itemsets," in *Proceedings of the 2003 SIAM international conference on data mining*. SIAM, 2003, pp. 59–70.

[6] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 1–12.

[7] S. Hanai, H. Nanba, and A. Nadamoto, "Clustering for closely similar recipes to extract spam recipes in user-generated recipe sites," in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*. ACM, 2015, p. 31.

[8] A. Jain, N. Rakhi, and G. Bagler, "Analysis of food pairing in regional cuisines of india," *PloS one*, vol. 10, no. 10, p. e0139539, 2015.

[9] A. Joshi and R. Kaur, "A review: Comparative study of various clustering techniques in data mining," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 3, pp. 55–57, 2013.

[10] T. Ozaki, X. Gao, and M. Mizutani, "Extraction of characteristic sets of ingredients and cooking actions on cuisine type," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2017, pp. 509–513.

[11] Y. Shidochi, T. Takahashi, I. Ide, and H. Murase, "Finding replaceable materials in cooking recipe texts considering characteristic cooking actions," in *Proceedings of the ACM multimedia 2009 workshop on Multimedia for cooking and eating activities*, 2009, pp. 9–14.

[12] N. Singh and G. Bagler, "Data-driven investigations of culinary patterns in traditional recipes across the world," *arXiv preprint arXiv:1803.04343*, 2018.

[13] R. Tuwani, N. Sahoo, N. Singh, and G. Bagler, "Computational models for the evolution of world cuisines," *arXiv preprint arXiv:1904.10138*, 2019.

[14] Y. van Pinxteren, G. Geleijnse, and P. Kamsteeg, "Deriving a recipe similarity measure for recommending healthful meals," in *Proceedings of the 16th international conference on Intelligent user interfaces*. ACM, 2011, pp. 105–114.

[15] L. Wang, Q. Li, N. Li, G. Dong, and Y. Yang, "Substructure similarity measurement in chinese recipes," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 979–988.

[16] S. Yokoi, K. Doman, T. Hirayama, I. Ide, D. Deguchi, and H. Murase, "Typicality analysis of the combination of ingredients in a cooking recipe for assisting the arrangement of ingredients," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2015, pp. 1–6.

Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020)
IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2

2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) 978-1-7281-4889-2/20/$31.00 ©2020 IEEE 10.1109/ICCMC48092.2020.ICCMC-00034

Hybrid Model for Lung Nodule Segmentation based on Support Vector Machine and k-Nearest Neighbor

Srishti Sharma
Dept. of Electrical Engineering
Delhi Technological University
New Delhi, India
srishtish21@gmail.com

Prasenjeet Fulzele
Dept. of Computer Science
Jaypee Institute of Information Technology, Uttar Pradesh, India
prasenjeet.fulzele7@gmail.com

Indu Sreedevi
Dept. of Electronics & Communication Engineering
Delhi Technological University
New Delhi, India
s.indu@dce.ac.in

ABSTRACT

The synergy of recently developed diagnostic radiology and machine learning algorithms has assured far reaching implications for the healthcare industry. At present, radiologists have access to top notch computer aided diagnostic (CAD) systems to create a consequence of the amplifying use and substantial applications of AI tools built right on the top of simple machine learning algorithms. This article proposes a model that extracts lung nodules from a 2 dimensional computed tomography (CT) slice by utilizing synthetic minority over-sampling technique (SMOTE) along with support vector machine (SVM) and k-nearest neighbor (K-NN) on a dataset of SPIE-AAPM Lung CT Challenge, 2015. Morphological transformations were performed on the 2D CT slices to achieve lung segmentation. Shape and textural features were retrieved into a vector to represent the region of interests (ROIs) from the lungs. Further, SMOTE was applied to resolve the issue of an imbalanced training data set which had very few samples of positive class in comparison with the samples of negative class. This ensured unbiased training of the classifiers and higher sensitivity towards the positive class. In the proposed work, two binary classifiers are combined in order to get an efficient model that exploited the individuality of both the classifiers. First, SVM and k-NN are trained separately on the balanced training dataset and then the outputs of both the classifiers are combined using simple sum rule to make the final prediction based on the collective scores for each data sample. Consequently, the resultant predictions depend on the collective performance of both classifiers for enhancing the overall efficiency of the model. The proposed hybrid model of SVM-k-NN outperforms the individual models with a sensitivity of 94.45% and G-Mean value of 94.19%. The model concentrates on accurately predicting the presence of a nodule and not for misclassifying a positive sample as it may lead to a huge loss to the patient.

CCS CONCEPTS

• Diagnostic radiology • computer aided diagnostic system (CAD) • machine learning

KEYWORDS

Support vector machine (SVM), segmentation, SMOTE, K-NN

1. INTRODUCTION

Lung Cancers are proclaimed to be the leading cause of deaths around the globe, as they've proven to be more fatal than other common forms of cancers. Specifically in Asian countries, where around 60% of the world's population is believed to reside, the lung cancer fact sheets demonstrates that the continent accounts for more than half of the total lung cancer cases reported on the planet. The diagnosis of a lung cancer is achieved through a computer aided diagnostic system. Computer aided diagnostic systems (CADs) have been making their way into the health care industry for a long period of time now and have witnessed a tremendous growth followed by a massive boost in the use automation worldwide in almost every field of research and development. CADs are employed for the screening of the thoracic region and to locate abnormalities if any present inside the patient's lungs [1].

In this paper pulmonary nodules found inside the lungs on a CT scan are taken up as the subject of concern which may later on be examined by the radiologists for further evaluation or to determine the staging of a cancer. CT scan is the most common and relevant imaging modality adopted for such kind of diagnosis. CT scans are X-ray images of a patient reconstructed using a CAD system that generates a high quality 3-D imaging with high resolution so that even structures of the order of a couple of millimeters could be spotted. Even on a CT scan it is sometimes hard to pin point and classify something suspicious within the lungs as a nodule and hence other screening technique such as a PET scan could be done to confirm the presence and delineate the characteristics of a nodule. This article aims at eliminating the need for a second opinion by resorting to another modality by improving the CADs performance. Pulmonary nodules generally found on chest X-ray or CT scan and could hint either a lung cancer or some deformity or lesion either caused by infection or inflammation or due to a previous trauma. In medical imaging the most crucial step for the physician is to identify any defect or abnormality without confusing it for any anatomical structure say soft tissues or blood vessels etc. due to their striking resemblance with each other when seen on the monitoring device. Therefore, nodule segmentation is vital in early diagnosis and prognosis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

In healthcare economy saving time equals saving lives, so indeed there's a solid need for automation in this field for not rushing into conclusions but only to assist the experts of the field be it physicians, pathologists or radiologists and to save precious time.

1.1 RELATED WORK

Recent developments in lung nodule extraction includes a U-net based image segmentation on the Kaggle data science bowl and LUNA 16 dataset, that proved to be an imperative step in detecting the candidates which actually had a tumor and then utilized 3D CNN for tumor classification [2]. In another case a two phase deep learning based model that envelops both nodule extraction and false positive reduction. In this case two networks namely the DNN and CNN were used together to contribute in reduction of false positives. This CNN nodule classifier attained a sensitivity of 82% [3]. A similar model employed a deep convolutional network on the ImageNet LSVRC 2010 dataset trained on over 1 million high-resolution images to be classified into 1000 unique classes [4]. Imaging studies involving different modalities, genetic algorithm (GA) and particle swarm optimization (PSO) was applied in order to improve the tumor detection [5]. An interesting combination of Support Vector Machine (SVM) and GA has shown to reach around its full potential accuracy of the classification process [6].Nodule detection by deploying image processing techniques such as feature extraction using deep neural networks abiding by a supervised learning [7] is one novel methodology of its kind. A unique computer aided diagnostic system (CAD) system was developed to detect nodule in the CAT scans of the thoracic region and was imposed on the famous lung image database consortium (LIDC) dataset. A CNN was trained by deploying the back propagation algorithm. The system produced decent results by performing with a sensitivity of around 78.9% [8].

2. METHODOLGY

The paper is implemented in two steps; the first step involves segmentation of the region of interest (ROI) from the whole 2-D CT scan image. In order to discard the falsely detected regions, certain features are extracted from the segmented regions. In the second step, machine learning model is used for the classification of segmented area into nodule and non-nodule region. The features extracted in the previous step are injected as input in the proposed machine learning model for classification.

2.1 Image Segmentation

In order to detect the nodule, the image used is a two dimensional CT slice from the 3D CT scan of the lungs. The lungs and nodules are segmented from the image using morphological operations [8]. In the event of false detection an offline trained classifier is used to reject that region.

Lung segmentation: The lungs are separated from the image using morphological transformations as depicted in Fig. 1(a-f). First the global threshold is calculated for each 2D slice separately represented in Fig. 1(a). The image is then converted into binary format as shown in Fig. 1(b). The image is eroded in order to remove the disconnected pixels or smaller regions as shown in Fig. 1(c). The image is further complemented as shown in Fig. 1(d). The edges are smoothened by using the close operation as shown in Fig. 1(e). The borders of the image are removed and any pixels with logical value 0 surrounded by value 1 are complemented by filling the image as shown in Fig. 1(f). The regions outside the lungs are removed and final masks are obtained as shown in Fig. 1(g).

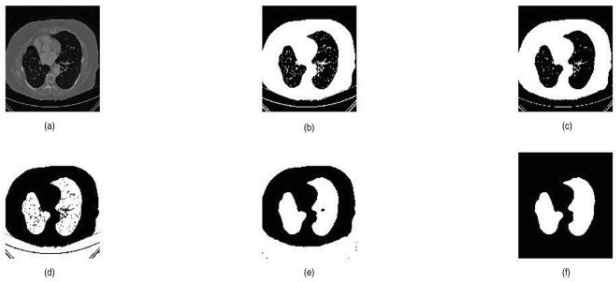


Fig. 1 (a)-(f) Lung segmentation

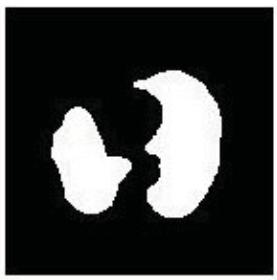


Fig. 1 (g) Segmented lungs



Fig. 1 (h) Segmented candidate nodules

Nodule Segmentation: The mask obtained from lung segmentation is used to extract the candidate nodules. The masked image is multiplied with the original 2D CT scan

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

image to obtain the ROIs which are represented in Fig. 1(h). These regions include some false detections which need to be discarded to accurately identify the nodule.

2.2 Feature Extraction

Certain features are extracted from the regions detected in the previous section in order to gather significant information for classification into nodule and non-nodule region. To characterize the regions suitably two type of features, shape and texture features are extracted [9]. Each segmented region's unique properties can be evaluated to depict the individuality of every region. Shape features that are extracted are area, circularity and eccentricity. The texture features are evaluated over the Grey Level Co-variance Matrix (GLCM) of an image [10]. These features represent the degree of heterogeneity, smoothness and contrast. In the proposed work the features used are energy, entropy, homogeneity, contrast, correlation, mean, variance and normalized inverse difference moment. These features contribute towards an accurate physical representation of segmented regions.

2.3 Dataset preparation

Amongst the ROIs detected, the number of nodules is limited as compared to other soft tissues and bronchioles present inside the lungs. This leads to the problem of imbalanced dataset which effects the identification of the minority class. To maintain a balanced proportion between the observations of the two classes sampling techniques are designed in order to over sample or under sample the dataset. The most widely used technique is the synthetic minority over-sampling technique (SMOTE) which involves the formation of similar data samples as that of the minority class to increase its number of observations and achieve a balance between the samples of both classes.

Synthetic minority over-sampling technique (SMOTE) is an imperative measure taken to circumvent over-fitting in a situation where the primary dataset is blended with exact copies of the minority instances. To achieve this a subset from the minority labels is taken to produce newer synthetic instances similar to the original ones. These newly produced synthetic instances and the original dataset can then be merged to be used as samples in training the classifiers. While by simply deploying both under and oversampling techniques could resolve the imbalanced samples in the dataset but the newly created sample may later turn out to be invalid in the process of creating a decision making interface and therefore only by using these algorithms does not necessarily uplift the classifiers accuracy in labeling the minority class.

SMOTE on the other hand works in a slightly different manner. The algorithm first maps out the distances between samples from both the training data and minority class samples and divides the samples into three groups namely – Border Samples, Safe samples and latent noise samples. To achieve this, distances between the training data and minority class samples is estimated. The Safe and Noise categorized samples can improve and disrupt the classifier's performance respectively. Finally a data point is chosen at random by the algorithm from the K-nearest neighbors of both safe sample and border sample categories.

2.4 Nodule Classification Model

After the preparation of a balanced dataset the samples are embedded into the proposed machine learning model for classification of the nodule resulting into a binary classifier. The model is implemented in two parts. The first part involves the training of individual SVM and k-NN classifiers. Further in the second step the scores from both of the classifiers are combined using simple sum rule to make the final resultant classification.

i) Support Vector Machine

Support vector machine (SVM) being a strong classifier, is a fairly popular machine learning algorithm. When it comes to handling binary classification problems it utilizes a certain mathematical function to its maximum potential. In almost every case SVM is capable to render highly accurate results with respect to a given dataset. SVM brings in supervised learning and connected learning algorithms together which are utilized for defining classes and analysis. Given a data input it maps the data points in to N-dimensional space and creates a clear demarcation about the class each of these points belong to. In binary classification, the given data in an input is segregated by a hyperplane which in turn classifies the input fed into a pair of labels and in this case whether or not a given input is a nodule. For the vectors that fall close to the defined boundaries of SVM are referred to as support vectors while the margins refer to the distance in between the hyper plane and support vectors [11].

ii) K- Nearest Neighbor

K-nearest neighbor predicts the values of instances based on the k closet training samples in the dataset [12]. For prediction of each instance, the most similar instance is searched for and the final prediction is made based on the k- closest instances present near the new instance. While dealing with the problem of classification k-NN works towards finding the class with the highest frequency to which the k-nearest neighbors belong. It is an instance based learning which goes through the whole dataset to make the decision for a single instance, thereby making it a lazy procedure.

The posterior probabilities from both the binary classifiers are added up which represent the score of each sample

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

belonging to the corresponding class. This helps in exploiting the advantage of each model over the other consequently enhancing the performance of the whole model. This results into a hybrid model of SVM and k-NN trained on a balanced dataset prepared using SMOTE.

3. IMPLEMENTATION

The training and evaluation of the proposed model is performed on the SPIE-AAPM Lung CT Challenge dataset which is publically available on the cancer imaging archive (TCIA) [13]. The images from the provided dataset are used to extract features in order to train the proposed hybrid model. The model is implemented in two stages where in first stage involves the training of individual classifiers and second stage consists of combining the results from the previously obtained classifiers to get the final prediction. The performance of the proposed model is evaluated based on the performance metrics accuracy, sensitivity and G-mean.

3.1 Dataset and Features

The SPIE-AAPM Lung CT Challenge dataset retrieved from the TCIA consists of 3D CT scans of 70 patients in DICOM format. Each scan comprises of 2D CT slices varying in number for every patient with a dimension of 512*512 pixels. The dataset is divided into two parts one of which is provided for calibration which includes data of 10 patients. The test set consists of CT scans of 60 patients. The nodule location for every patient in the calibration set is provided beforehand with the corresponding instance number of the 2D slice which is considered as the Nodule Centre Image. The instance number of each image is numbered at (0020, 0013) in the metadata. The nodule center image of each patient is used to segment the nodule of the patient resulting into data of only 10 nodules for training of the proposed model. The rest of the nodule candidates dominate the actual nodules in number.

3.2 Nodule Segmentation

In order to extract the candidate nodules from 3D CT scans the lungs were first segmented from the 2D image slice using thresholding and morphological transformations as shown in Fig. 1. The ROIs were detected by simply multiplying the segmented lungs with the original 2D CT slice. One of the ROIs of a particular scan was the nodule for the corresponding patient. In order to represent the ROIs in numerical form, certain features are extracted from these regions. Three shape features and eight texture features are extracted. The shape features extracted are area, circularity and eccentricity and the texture features used for the proposed work are energy, entropy, homogeneity, contrast, correlation, mean, variance, and normalized inverse difference moment. The final feature vector is 1*11 dimensional and it is fed as input to the classifiers later on as shown in Fig. 2. The dataset was normalized in order to avoid unequal distribution of weights amongst the features and achieve fast convergence. The problem of imbalanced dataset is overcome by using the SMOTE sampling technique which synthetically generated samples for the minority class so as to maintain a proportion between the samples of the two classes. Before applying SMOTE the number of samples for Nodule class was 10 which was the number of patients of which the calibration set consisted. On the other hand, the number of non-nodule class samples was 373. This clearly would have resulted into a biased model which detected the non-nodular regions much more accurately than the nodular regions. After over sampling of the dataset using SMOTE the number of observations for nodule class was increased to 373 resulting into a training dataset of 746 samples with equal distribution of samples between the two classes. Now to identify the nodule correctly and discard the other segmented regions a machine learning model was trained which classified the regions as a nodule or a non-nodule.

3.3 Classifiers

The dataset prepared was in normalized form and given as input to the two separate binary classifiers. The results of both the classifiers which were in the form of posterior probabilities were combined using the simple sum rule. The hyper parameters of the classifiers were optimized using Random Grid Search.

i. A binary classifier was built using SVM on a linear kernel with Iterative Single Data Algorithm [14] as the optimization technique. The posterior probabilities are calculated for each sample that represent the extent to which a sample belongs to a particular class.
ii. Another binary classifier was trained using k-Nearest Neighbor with number of neighbors as 15 and exhaustive nearest neighbor search method. The distance metric used was Euclidean.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

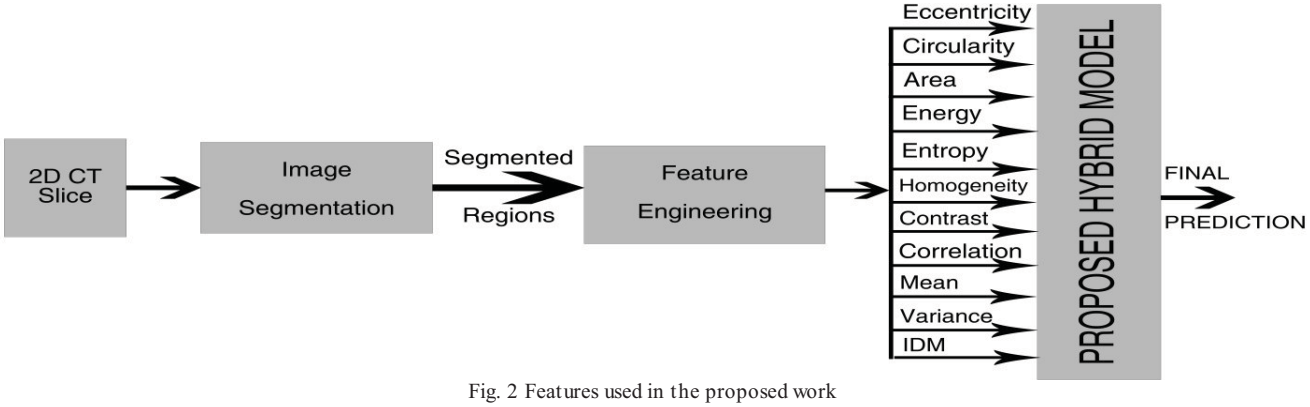


Fig. 2 Features used in the proposed work

The posterior probabilities which were the scores of each class were calculated for this classifier as well.

Since the performance of both the classifiers indicated satisfactory results on non-overlapping samples, combining the results of the two for the final prediction exhibited a significant improvement over the performance of both the individual classifiers. The structure of the final proposed hybrid model is shown in Fig. 3. The performance measures used to evaluate the models are accuracy, sensitivity, specificity and G-mean.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{TN + FP} \quad (3)$$

$$\text{G} - \text{Mean} = \sqrt{Sensitivity * Specificity} \quad (4)$$

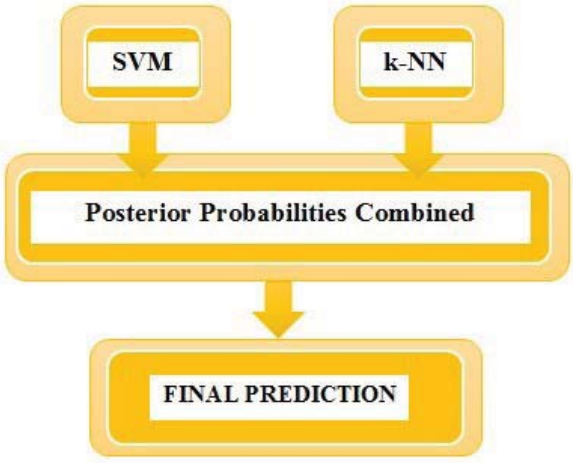


Fig. 3 Structure of Proposed Hybrid Model

4. RESULTS

The training set and testing set consisted of 746 and 2959 samples respectively. The observations of the training set were equally distributed amongst the two classes. Random Grid Search was used for hyper parameter tuning. Imbalanced dataset leaded to a biased classifier with 0% sensitivity as the samples for the nodule class were not enough in number. There were no true positives or false negatives and all samples were predicted as a non-nodular region. But the classifiers exhibited satisfactory performances with the balanced dataset generated using SMOTE. Further the proposed hybrid model of SVM-k-NN outperformed both the classifiers when trained individually with a sensitivity of 94.45%. The main objective of the current work is to detect the nodule accurately. Increased number of false positives is not harmful since a non-nodule region detected as a nodular one would lead to the examination of the region by the radiologist and later on concluding that it is not a nodule. On the other hand, increased number of false negatives is an intolerable output since in this case a nodular region is already discarded and would not be examined by the radiologist which is a major loss to the patient. The proposed hybrid model exhibits the desired performance and is clearly an improvement over the individual classifiers as shown in Fig. 4. The confusion matrix is shown in Table 1 which showcases the performance of the proposed hybrid model.

The comparison of the proposed hybrid model with the individual classifiers is shown in Table 2 on the basis of performances metrics described earlier. Due to the summation of individual scores from each of the classifier the performance of the final model was evidently enhanced by taking in account the advantages of both the classifiers.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

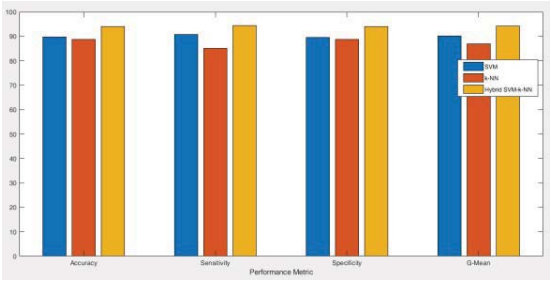


Fig. 4 Performance Comparison of all models

TABLE I CONFUSION MATRIX OF PROPOSED MODEL

	Nodule	Non-Nodule	Total
Nodule	**51** (TP)	3 (FN)	54
Non-Nodule	176 (FP)	**2729** (TN)	2905
Total	227	2732	2959

TABLE II PERFORMANCE COMPARISON OF MODELS

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	G-Mean (%)
SVM	89.66	90.74	89.64	90.19
k-NN	88.64	85.19	88.71	86.93
Proposed	**93.95**	**94.45**	**93.94**	**94.19**

5. CONCLUSION

The proposed hybrid model of SVM and k-NN using the SMOTE sampling technique out performs both the individual classifiers and models with imbalanced dataset clearly exhibiting an accuracy of 93.95% and a sensitivity of 94.45%. The sensitivity plays a major role in evaluating a models performance since it depicts the efficiency of the model in accurately detecting a nodule and not discarding it falsely. Segmentation of the nodules using morphological operations makes it easier for a machine to learn about a nodule since it enables us to extract the features of only the relevant regions, the ROIs. Reducing the problem of identification to only a few regions of interest makes it a less time consuming and simpler task for the machine to perform. The identification of the nodule is then efficiently automated by the proposed classifier. The proposed model is evidently an improvement over the classifiers separately trained and evaluated individually.

6. ACKNOWLEDGMENTS

The dataset was retrieved from The Cancer Image Archive (TCIA) which consists of a large number of archives of medical images of cancer and is accessible for public download. The dataset used was uploaded for the LUNGx Challenge as part of the 2015 SPIE Medical Imaging Conference, SPIE – with the support of American Association of Physicists in Medicine (AAPM) and the National Cancer Institute (NCI). This challenge was started for quantitative image analysis to allow the participants to design algorithms for classification of nodule into benign or malignant and the ground truth of the calibration set was also provided. This resource was put to the benefit of the proposed work and is publically available at https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM+Lung+CT+Challenge.

REFERENCES

[1] W. CHOI, T. Choi, Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach, Entropy 2013, 15, pp. 507-523.

[2] W. Alakwaa, M. Nassef, A.Bad, Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN), International Journal of Advanced Computer Science and Applications(IJACSA), Volume 8 Issue 8, 2017.

[3] J. Tan,Y. Huo,Z. Lian, L.LI, Proceedings of the Symposium on Modeling and Simulation in Medicine, Article No.8,2017

[4] A. Krizhevsky,I.Sutskever, G. Hinton„NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems – vol. 1, pp. 1097-1105.

[5] A. Bhardwaj, A. Tiwari, "Breast cancer diagnosis using genetically optimized neural network model," Expert Systems with Applications,
vol. 42, pp. 4611-4620, June 2015.

[6] N. Azizi, N. Zemmal, M. Sellami, and N. Farah, "A new hybrid method combining genetic algorithm and support vector machine classifier: Application to CAD system for mammogram images," in Multimedia Computing and Systems (ICMCS), 2014 International Conference on, Marrakech, , pp. 415-420, April 2014

[7] Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, and E. I. Chang, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2014, pp. 1626–1630.

[8] R. Golan,C. Jacob, J. Denzinger, "Lung Nodule Detection in CT Images using Deep Convolutional Neural Networks", in Proceedings of International Joint Conference on Neural Networks (IJCNN), Canada, 2016, pp. 243-250.

[9] Y. Sui, Y. Wei, D. Zhao, "Computer-Aided Lung Nodule Recognition by SVM Classifier Based on Combination of Random Undersampling and SMOTE", in Computational and Mathematical Methods in Medicine Volume 2015, Article ID 368674, pp. 1-13.

[10] R. A. Thomas, S. S. Kumar, "Automatic detection of lung nodules using Classifiers", International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), India, 2014, pp. 705-710.

[11] E. R. Gonzalez, V. Ponomaryov, "Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM", in Proceedings of 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW), Ukraine, 2016, pp. 1-4.

[12] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," IEEE Trans. Syst. Man. Cybern., Nov. 1973, vol. 3, no. 6, pp. 610–621.

[13] Duraipandian, M. "Performance Evaluation of Routing Algorithm for MANET based on the Machine Learning Techniques." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 01 (2019): 25-38.

[14] V. Kecman, T.M. Huang, M. Vogt. "Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance." In Support Vector Machines: Theory and Applications, Berlin, 2005, pp. 255–274.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:47 UTC from IEEE Xplore. Restrictions apply.

Microsystem Technologies
https://doi.org/10.1007/s00542-020-04845-2

TECHNICAL PAPER

Investigation of electrical/analog performance and reliability of gate metal and source pocket engineered DG-TFET

Jaya Madan[1] · Rahul Pandey[1] · Rajnish Sharma[1] · Rishu Chaujar[2]

Received: 14 March 2019 / Accepted: 7 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract
Although Tunnel Field Effect Transistors (TFET) offer low leakage current and allow good scalability, but they suffer from low ON-current. Present research paper focuses on the designing of an industry ready TFET that can operate efficiently at nano-scales allowing the fabrication of a smaller and more energy efficient overall device. To achieve the same, the Gate Metal Engineering (GME) and n^+ Source Pocket (SP) schemes have been integrated on Double-Gate TFET (DG-TFET). For a distinct examination of the merits of GME and SP, 4 types of TFET architectures, namely the DG-TFET, GME-DG-TFET, SP-DG-TFET and GME-SP-DG-TFET (which amalgamates the merits of both GME and SP engineering) are investigated. All these device architectures are examined in terms of their electrical characteristics and analog parameters. GME-SP-DG-TFET (being superior among the four) is further analyzed under the presence of interface fixed charges (FC) of different polarity to affirm its reliability. Investigations of this device structure reveal that the positive FC are more noxious for the device as compared to negative FC. However, the enhanced flatband voltage under the negative FC translates to the high gate bias requirement to turn the device ON. Observance of only marginal variations in the parasitic capacitances and efficiency of the GME-SP-DG-TFET device in the presence of FC (both positive/negative) signifies its immunity and promises enhanced reliability. A remarkable improvement of the I_{ON}/I_{OFF} by two order of magnitude (from 10^{10} to 10^{12}) and a decrease in V_{th} by 27.71% observed for GME-SP-DG-TFET as compared to DG-TFET along with the improved reliability makes the former an efficient candidate for low power analog/RF applications.

1 Introduction

Over the years, driven by industry for "smarter" products, engineers have continued to pack an ever-increasing number of transistors onto a single chip. In this process, engineers have had to not only develop novel fabrication techniques, but also make MOSFETs smaller. At the same time, the demand to make CMOS devices smaller has led to a reduction in the physical size of the chip. This two-fold crunch necessitates the design of transistors with dimensions well within nano-scales. Shrinking of MOSFETs beyond a certain point, however, gives rise to certain phenomena referred to as Short Channel Effects (SCEs) that degrade the performance of the transistor (Madan et al. 2017b; Villalon et al. 2014). There is also an increase in the leakage current through the device, which adversely affects the use of such transistors in low-power applications. As these shortcomings of the MOSFET came to the foray, researchers began to look at new paradigms of transistor construction that are suitable for nano-scale operation. Among the devices proposed, the Tunnel FET (TFET) has emerged as a promising device attracting much research. This success of TFETs can be attributed to the fact that the primary source of drive current in TFETs is quantum mechanical Band-to-Band Tunneling (BTBT), which, being a quantum phenomenon, is prevalent at nano-scales (Biswal et al. 2017; Kumar and Raman 2019). This is contrasting to thermal carrier injection in MOSFETs which

✉ Rahul Pandey
rahul.pandey@chitkara.edu.in

✉ Rajnish Sharma
rajnish.sharma@chitkarauniversity.edu.in

Jaya Madan
jaya.madan@chitkara.edu.in

Rishu Chaujar
chaujar.rishu@dtu.ac.in

[1] VLSI Centre of Excellence, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

[2] Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 18 April 2020

itself imposes certain physical limitations on the operation of the transistor. The construction of TFETs is similar to that of a MOSFET. A basic TFET is a four-region device with a source, channel, gate and drain. The source and drain are doped opposite to each other. TFETs work on the principle of BTBT of charged particles through the potential barrier at the source-channel interface. This mechanism makes the operation of TFETs at low supply voltages possible leading to lower static and dynamic power dissipation (Joshi et al. 2020). The key merits of the TFET are the low off current (I_{OFF}), steeper subthreshold slope and process compatibility with the standard CMOS process flow (Choi et al. 2007; Nirschl et al. 2006).

As promising as the TFET may seem as a transistor candidate, it is far from being employable in modern electronics due to inherent complications in the device and the lack of focused R&D in the direction. A significant roadblock in such use of TFETs is the lower ON-state current (I_{ON}) and considerably large miller capacitance as compared to MOSFETs because of inefficient BTBT. Several researchers have reported various device/material engineering schemes to overcome the challenges of TFET. To reduce the miller capacitance and simultaneously to enhance the I_{ON}, Lu et al. explored a drain current and capacitance model of hetero gate dielectric TFET (Lu et al. 2018). The tunneling current can be enhanced by greater overlapping of the valence band on the source side and conduction band on the channel side (Wei et al. 2018). Kim et al., have fabricated a DG TFET with vertical channel inserted between lightly doped Si and reported a SS of 17 mV/decade with I_{ON}/I_{OFF} ratio of 10^4 (Kim et al. 2019). Li et al., have also fabricated a vertical TFET with a p-type SiGe pocket near the source side and reported I_{OFF} of 100 pA/μm (Li and Woo 2020). Besides, currently, the temperature dependence of analog performance, linearity and harmonic distortion for a Ge-Source Tunnel FET has been investigated by Datta et al. (2020). In their work, they have incorporated the source bandgap engineering to conquer the challenges of TFET, and have also reported the temperature robustness comprehensively. Additionally, Joshi et al. have proposed an extended source DG TFET intending to enhance the analog/RF performance of TFET. Their work reports threshold voltage of 0.42 V, subthreshold swing of 12.24 mV/decade, I_{ON}/I_{OFF} of 2.57×10^{12}, cut off frequency 37.7 GHz, and gain-bandwidth product of 3.4 GHz (Joshi et al. 2020). Among all, the source pocket in the channel near source side is the most effective engineering to enhance the band bending at the tunneling junction in TFET (Nagavarapu et al. 2008). The source pocket is a region of highly doped semiconductor between the source and channel that causes abrupt band bending at the source-channel interface. As a result, the energy levels of the conduction band of the channel are pulled closer to those in the valence band of the source, increasing the probability of electron tunneling through the device (Madan and Chaujar 2017b). The source pocket, along with gate stacking and gate underlapping engineering has been amalgamated on cylindrical gate TFET by Singh et al., to improve the electrical/RF performance of TFET (Singh et al. 2020). However, in this work, an n^+ source-pocket is used for this purpose, resulting in an n^+ Source Pocket TFET (SP-TFET). The n^+ pocket causes improvement of I_{ON}, reduction in V_{th} and enhancement of the current switching ratio (Madan and Chaujar 2017a).

Another bottleneck of TFET is caused by the asymmetrical source/drain doping, which results in severe ambipolar conductivity, allowing TFETs to conduct current under both positive/negative gate bias (Wang et al. 2004). Solution to this problem is to engineer the gate in such a fashion that the band bending must be smaller (larger tunneling barrier) near the drain side. The large tunneling width at the drain channel junction makes it more difficult for electrons to tunnel at a negative gate bias, subsequently lowering the ambipolar current. To accomplish this, an engineering technique referred to as Gate Metal Engineering (GME) has been reported (Cui et al. 2011; Madan et al. 2017b). in GME, a metal with low/high value of work function is used near the source/drain side. The lower work function of metal reduces the flat band voltage (V_{fb}=-$\phi_M - \phi_S$) and thereby enhances the effective gate bias ($V_{eff} = V_{gs} - V_{fb}$) at the tunneling junction. The enhanced V_{eff} increases the I_{ON} (due to reduction in tunneling barrier width at the source-channel interface) and also reduces the V_{th}. Likewise, the metal towards the drain side with comparatively higher work function controls the I_{OFF} and ambipolar conduction by broadening the barrier width at the drain-channel interface (Madan et al. 2017a).

Generally, in most of the research work, the Si/SiO_2 interface is treated as ideal (with no defects), however during the fabrication (owing to the stress and radiation induced damages) the interface is caused by mobile/fixed trapped charges. These charges effectively act as positive/negative fixed charges (FC) depending on their energy levels in the semiconductor bandgap. If a defect state is neutral and becomes positive/negative after donating/accepting an electron it is considered as a donor/acceptor state or a positive/negative FC (Madan and Chaujar 2017b, 2018; Qiu et al. 2014). The modification caused by the FC in the characteristics of the device will affect the functioning of the integrated circuit. Thus, to ensure the immunity of device against FC it is essential to investigate the characteristics of device under the presence of FC. Furthermore, GME confers another benefit on the TFET that is not immediately visible in the utopian scenario presented by simulation environments. It has been earlier seen that the application of the GME scheme makes the

transistor more immune to the presence of Interface Fixed Charges (IFCs) at the SiO_2–Si interface at the tunneling interface (Madan et al. 2017c). This improved immunity against IFCs directly translates to improved reliability of the device. Often the reliability of transistors is simply associated with certain physical phenomena like electro-migration that play a role only after the device comes into operation. The appearance of IFCs at the SiO_2–Si interface begins right at the time of fabrication and adversely impacts device reliability by creating intermediate energy levels that serve as transient states for tunneling electrons (Cao et al. 2011). Improved fabrication techniques have led to a significant reduction in the density of IFCs, but their elimination appears to be a physical impossibility. In any realistic device, it becomes essential to minimize the impact of FCs on the performance of the transistor by means of device design and engineering (Sant et al. 2016). Therefore, much like figuratively hitting two birds with one arrow, the GME scheme incorporated in this work helps to control both drain current and device reliability.

This work attempts to remedy the aforementioned shortcomings of TFETs without depriving the device of the improvements it had achieved over the other transistor candidates, often even improving upon its overall performance. The primary motivation behind this work is to design a transistor that is an inch closer to an industry ready transistor candidate and can simultaneously operate efficiently at nano-scales allowing the fabrication of smaller, more energy efficient devices. The paper is ordered as: Section II describes the device structure, simulation methodology and the calibration of our simulation setup (models), section III deals with results obtained by numerical simulations, and the explanations thereof. A number of performance parameters and industry standard FOMs have been used to analyze the viability of the use of the proposed devices for practical applications. Broadly, the result section is divided into two subsections: the first analyzing the electrical/analog performance of DG-TFET by integrating SP and GME engineering schemes, and the second investigating the device reliability in terms of FC. Finally, conclusions drawn in section IV indicate that the proposed Gate Material Engineered Source Pocket DG TFET (GME-SP-DG-TFET) is a viable device for low power, high-frequency operations.

2 Device architecture and simulation methodology

In this work, four different architectures of double gate Tunnel-FET (DG-TFET) have been considered for performance comparison. The four different structures are Double Gate TFET (DG-TFET), Gate Material Engineered DG TFET (GME-DG-TFET), n^+ Source Pocket DG TFET (SP-DG-TFET) and Gate Material Engineered Source Pocket DG TFET (GME-SP-DG-TFET). These devices are referred to as D1, D2, D3, and D4 respectively in this paper, and their schematic architectures are shown in Fig. 1a–d. In GME-DG-TFET, i.e., D2, both the gate metals (top and bottom) are engineered with dual material, i.e., with different metal work functions. Precisely, the GME scheme incorporates the lower metal work function near source side (tunneling gate) and comparatively higher metal work function near the drain side (auxiliary gate). Additionally, the SP-DG-TFET, i.e., D3 implants, a n^+ pocket near source side in the channel. Subsequently, the architecture that integrates the merits of GME and SP, i.e., GME-SP-DG-TFET referred as D4 comprises both, gate material engineering and also n^+ source pocket engineering. The source pocket width and doping concentration, in the n^+ source pocket engineering, are optimized for optimum I_{ON}/I_{OFF}, to 4 nm and 2×10^{19} cm^{-3} respectively (Madan and Chaujar 2017b). Further, the work function for tunneling gate (gate near the source) is 4.1 eV and that for the auxiliary gate is 4.3 eV in case of GME, again tuned to achieve an optimized I_{ON}/I_{OFF} ratio (Shekhar et al. 2018). Additionally, for GME, the tunneling and the auxiliary metal gate length is 10 nm and 40 nm respectively. Furthermore, source (p^+) and drain doping (n^+) has been kept 1×10^{20} cm^{-3} and 5×10^{18} cm^{-3} respectively, for each of the proposed structures (D1 to D4). The thickness of the oxide (SiO_2) layer is 2 nm for all the devices.

Further, to account the influence of FC, equivalent positive/negative charges are considered at the Si/SiO_2 interface in the simulated device architecture. The value of FC density, i.e., N_f is selected on the basis of several experimental reported data that incorporates process damage and radiation damage (Madan and Chaujar 2016; Pala et al. 2012; Sant et al. 2016). The literature shows that the FC density varies in the range of 10^{11}–10^{13} cm^{-2} (Jiao 2009; Lho and Kim 2005; Poindexter 1989; Trabzon and Awadelkarim 1998). Thereby, to investigate the influence of FC, the FC of different polarity (positive/negative) is considered in the channel region only near the source side (up to 10 nm from source side), as the FC are generated majorly at the region where there is a high electric field (Madan and Chaujar 2017b, 2018). Additionally, in the simulation setup, the INTERFACE statement is used to specify the FC density and position (at the Si/SiO_2 interface) of the FC.

All the simulations have been carried out using the TCAD SILVACO ATLAS device simulator (SILVACO 2011). The models incorporated in simulations are nonlocal BTBT model, bandgap narrowing model (BGN), Shockley–Read–Hall (SRH) recombination model, concentration and field dependent mobility model (CONMOB). Further,

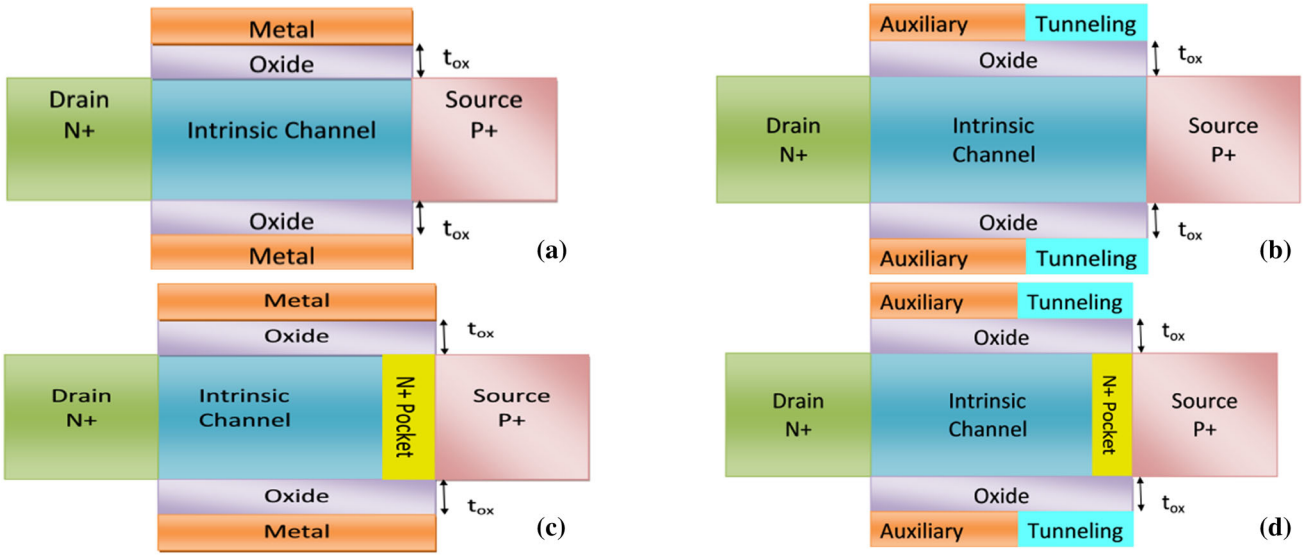


Fig. 1 2D device structure used in simulation **a** double gate TFET (DG-TFET) D1, **b** gate metal engineered double gate TFET (GME-DG-TFET) D2, **c** source pocket double gate TFET (SP-DG-TFET) D3, and **d** GME-SP-DG-TFET, D4

the Fermi–Dirac statistics and the conventional drift–diffusion model have been used. To perform the realistic study, models used during simulations are calibrated with experimentally reported work (Tura et al. 2011), i.e., a vertical DG-PNPN-TFET. To validate the simulation models the parameters of the nonlocal BTBT model (i.e., the prime essential model for TFET) is calibrated. In order to fine match the transfer characteristics of the simulated data with the experimentally reported data (Tura et al. 2011), the electron and hole effective mass are tuned to $0.22m_0$ and $0.51m_0$ from their defaults values of $0.322m_0$ and $0.549m_0$ respectively; where m_0 indicates the electron rest mass. Figure 2 depicts the calibrated transfer characteristics of our simulated setup with the published

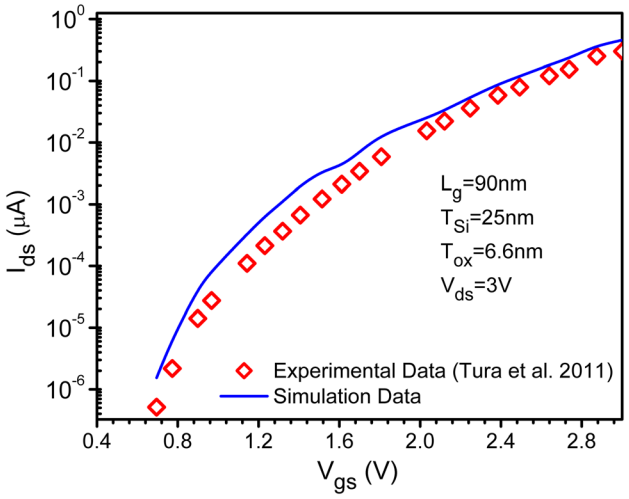


Fig. 2 Transfer characteristics calibration with experimental data reported in Tura et al. (2011)

experimental work of (Tura et al. 2011) and thereby validates the simulation setup.

The proposed device, i.e., Gate metal and source pocket engineered DG-TFET, can be realized experimentally. However, none of the literature reports the fabrication of the discussed device solely. Therefore, the proposed device can be realized as per the techniques mentioned in the literature. The n^+ source pocket DG TFET can be fabricated as discussed by Tura et al., where authors utilized molecular beam epitaxial (MBE) growth with a low thermal budget vertical process flow (Tura et al. 2011). After MBE, the hard mask of nitride is deposited and patterned for silicon wall etching to define the vertical channel, followed by gate oxidation using plasma-assisted oxidation. After that, the poly gate is deposited and etched along with the removal of the nitride spacer. Thereafter, an isolation oxide is deposited and planarized, followed by the deposition and patterning of the polysilicon gate metal. Besides the formation of n^+ source pocket in the proposed device i.e., Gate metal and source pocket engineered DG-TFET, a hetero material gate has to be deposited. Thus, to process the gate metal engineering (or to deposit the auxiliary and tunneling gate), the metal interdiffusion technique (Polishchuk et al. 2001), or the fully silicide metal gate structure (Liu et al. 2005), or metal wet etching process (Zhang et al. 2005), or the laterally joined two metal gate structure (Long et al. 1999) or the chemical mechanical processing technique (Guillaumot 2002) can be used.

3 Results and discussion

3.1 Device comparison: electrical/analog performance

In this subsection, the electrical and analog performance of the devices under considerations labeled as D1, D2, D3, and D4 are analyzed. To examine the electrical/analog performance of the devices, surface potential, energy bands, electric field, transfer characteristics, output characteristics, threshold voltage, Subthreshold swing, current switching ratio, and device efficiency are investigated.

The surface potential of the four devices under consideration, i.e., D1, D2, D3 and D4 in both ON and OFF states are depicted in Fig. 3a. It is evident that the surface potential of all the devices (in the channel region) is higher in the ON state. Further, the D4 has a higher potential at the tunneling junction followed by D3, D2, and D1. The higher potential of D2, i.e., GME-DG-TFET is attributed to the existence of tunneling gate metal (lower work function metal) near the tunneling junction. This lower workfunction of the tunneling gate reduces the V_{fb} and thereby enhances the V_{eff} under it, i.e. at the tunneling junction (Saurabh and Kumar 2011). The increase in V_{eff} results in the higher surface potential of D2 as compared to D1. Moreover, for D3, i.e., SP-DG-TFET, the implantation of n^+ source pocket in the channel near source side results in thin depletion layer at the source pocket junction (Nagavarapu et al. 2008). Thereby, the surface potential increases for D3 in comparison with D1. Furthermore, the integration of GME and SP collectively, i.e., GME-SP-DG-TFET (D4) shows superior surface potential among the 4 devices. The superior surface potential of D4 is mainly offered by the aggregated effect of GME and SP.

Figure 3b represent the energy of the electron along the length of the channel when the device is in the 'ON' state for all the four devices. It is analyzed from the plot shown in Fig. 3b that all the devices under consideration have narrow tunneling barrier. The graph also depicts that the band bending is steepest in case of D4, followed by D3, D2 and lastly D1. The steeper band bending in D2 is offered by the tunneling metal gate that is of low metal work function. The low work function of tunneling gate enhances the surface potential and thereby results in steeper band bending as depicted in Fig. 3b. Further, in D3, it is the n^+ SP that allows steeper band bending, between the source and source-pocket. Subsequently, in D4, there is the collective effect of GME and SP, that translates to the steepest band-bending as compared to all other devices.

Figure 4a, b depicts the electric field for the four considered devices D1, D2, D3 and D4 along the length of the channel in the OFF (@V_{gs} = 0.0 V and V_{ds} = 1.0 V) and ON-state (@V_{gs} = 1.2 V and V_{ds} = 1.0 V) respectively. It is clearly illustrated in Fig. 4a and b, that for each device at the tunneling junction, the electric field in the ON-state is higher than the OFF state. The high value of an electric field in the ON state is due to the contribution of the external electric field coming into existence due to gate bias that adds to the internal existing electric field (across the source-channel interface for D1, D2 or source-pocket interface for D3, D4). As such it results in a more strong electric field $\sim$ 3.5 MV/cm, whereas, in the absence of the gate bias, the diffusion of charged particles due to the concentration gradient results in the comparatively low internal electric field. Furthermore, in the ON as well as OFF state, the electric field is maximum in case of device D4, followed by D3, D2, and D1. The highest electric field in D4 is due to the steeper bending (discussed in Fig. 3b) offered by the collective effect of SP and the low work function of the tunneling gate. Moreover, it is also illustrated that the electric field at the drain channel junction is lower in case of ON state as compared to OFF state. The

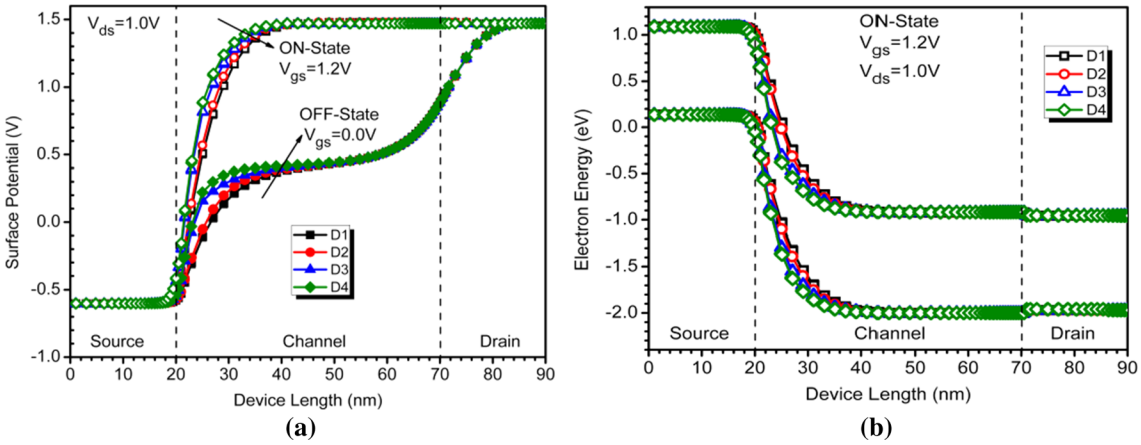


Fig. 3 **a** Surface Potential along the device length at V_{gs} = 0.0 V (OFF-state) and V_{gs} = 1.2 V (ON-state) at V_{ds} = 1.0 V. **b** Energy band diagram along the device length at V_{gs} = 1.2 V (ON-state) at V_{ds} = 1.0 V for all the devices under consideration

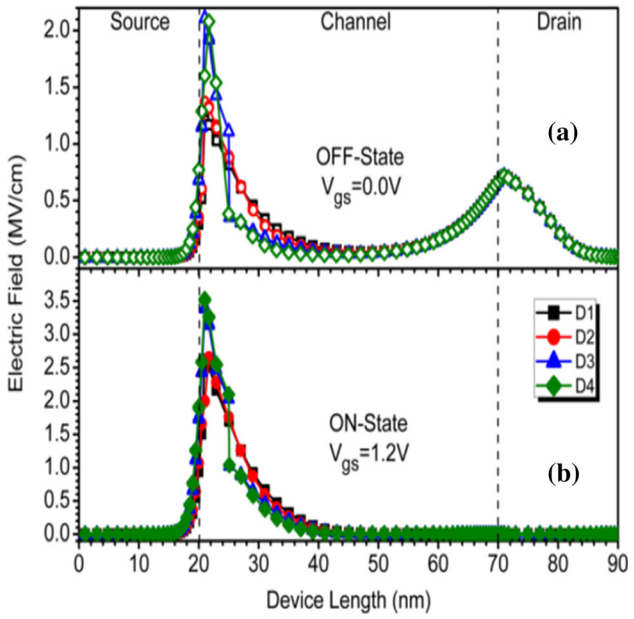
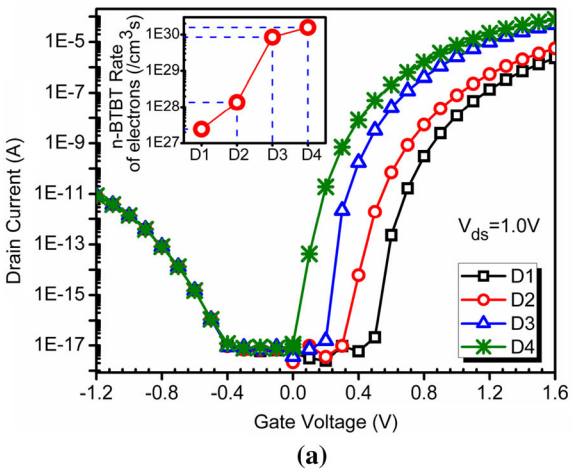


Fig. 4 Electric field along the device length at **a** V_{gs} = 0.0 V (OFF-state) and **b** V_{gs} = 1.2 V (ON-state) at V_{ds} = 1.0 V for all the devices under consideration

lower electric field at the drain-channel interface in ON state is attributed to the reduced drain channel potential barrier with rising V_{gs} (Boucart and Ionescu 2007).

The number of electrons tunneling at the source channel (D1/D2) or the source pocket junction (D3/D4) per cubic volume per second is plotted in the Inset of Fig. 5a for all the devices under consideration in the ON-state. It is clearly presented that the BTBT rate of electrons for D4 (10^{30}/cm^3 s) is remarkably higher than D1 (10^{27}/cm^3 s). The higher BTBT rate of electrons of D4 is facilitated by the narrower barrier width (offered by the source pocket and lower metal work function of tunneling gate) that permits large number of electrons to tunnel at the tunneling junction. Further, it must be noted that the BTBT rate of D2 and D3 is lower than D4 by many orders. Figure 5a also represents drain current as a function of gate voltage for all the four n-channel devices under consideration keeping the drain-to-source voltage at 1.0 V. Primarily, as the gate voltage is gradually increased the band bending at the tunneling junction leads to narrowing of tunneling barrier width. This narrowing of the tunneling barrier (less than 5 nm) eventually enhances the drain current.

Furthermore, the graph depicts that at a particular value of positive gate bias, device D4 has the highest current followed by D3, D2 and then D1. The higher drain current of D4 is due to higher tunneling rate of electrons as depicted in the Inset of Fig. 5a. It is evaluated that the I_{ON} of D1 increases from 10^{-7} A to 10^{-5} A with the incorporation of GME and source pocket, i.e., for D4. Moreover, when the gate is negatively biased, there is band bending at the drain/channel interface causing ambipolar conduction. Furthermore, the ambipolar current is same for all the four considered devices and is also tabulated in Table 1. The similar ambipolar conduction is attributed to the fact that the drain/channel junction (that principally controls the ambipolar conduction) is same for all the devices (Wang et al. 2004). Since ambipolar current depends upon the effect of negative gate bias on the drain-channel interface (for n-TFET) which is same for all the four devices, this leads to more constant values of ambipolar current. Moreover, comparing with the existing literature, where a III–V staggered hetero-junction nanowire (NW) TFET is used to improve the ON-state current. It is investigated that the I_{ON}/I_{OFF} ratio of the device reported by Biswal et al. is only 10^4 (for Si) whereas, in our work (by integrating the



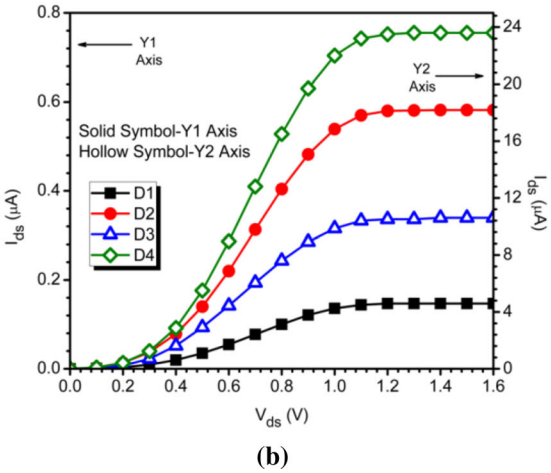


Fig. 5 **a** Transfer characteristics of all the devices under concern, i.e., D1, D2, D3 and D4 at V_{ds} = 1.0 V. INSET: Band to band tunneling rate of electrons (/cm^3 s) at the tunneling junction for D1, D2, D3 and D4 in ON-state @V_{gs} = 1.2 V and V_{ds} = 1.0 V. **b** Output characteristics of D1, D2, D3, and D4 at V_{gs} = 1.2 V

Table 1 V_{th}, SS, I_{ON}/I_{OFF} and I_{AMB} of D1, D2, D3, and D4

Parameters device	V_{th} (V)	SS (mV/decade)	I_{ON}/I_{OFF}	I_{AMB} (A)
D1	0.86	24.74	2.25×10^{10}	8.53×10^{-12}
D2	0.82	35.64	7.66×10^{10}	8.53×10^{-12}
D3	0.71	19.39	1.20×10^{12}	8.49×10^{-12}
D4	0.62	27.96	2.63×10^{12}	8.49×10^{-12}

device engineerings) the I_{ON}/I_{OFF} ratio of 10^{12} is achieved (Biswal et al. 2017). Figure 5b represents the output characteristics of all the devices at V_{gs} = 1.2 V. To highlight the significant changes in I_{ds}, the values of D1 and D2 have been shown on Y1 axis while the same has been shown on Y2 axis for D3 and D4. The proposed D4 provides a drain current of the order of 10^{-5} A in comparison to 10^{-7} A being provided by D1 when both the devices have been put into similar operating conditions. This significantly large drain current of D4 (almost 2 order of magnitude greater than other conventional techniques) may have far-reaching positive consequences to circumvent the challenges faced by TFET.

Important electrical parameters, i.e., the threshold voltage (V_{th}), current switching ratio (I_{ON}/I_{OFF}) and the sub-threshold swing (SS), extracted from the transfer characteristics, have been tabulated in Table 1 for all the devices under considerations. In this work, V_{th} is evaluated as the V_{gs} at $I_{ds} = 10^{-7}$ A, i.e., by the constant current method. It is clearly observed that the V_{th} of D4 is minimum. This minimum V_{th} of D4 is obtained by the higher tunneling rate that switches ON the device at lower gate bias. Further, the D3 provides the minimum SS among all the devices. However, the SS of D4 is still appreciably lower than the fundamental limit of MOSFET on the SS. Since I_{OFF} is almost same for all the four devices, the current ratio therefore predominantly depends on the I_{ON}. I_{ON} is found to be maximum in case of D4, leading to maximum current switching ratio in the case of D4, followed by D3, D2, and D1.

To provide an insight for the improvement of analog performance by a device in the sub-threshold region, device efficiency also known as transconductance generation factor (TGF) is one of the key figure of merit (FOMs) (Mallik and Chattopadhyay 2012). This device efficiency or TGF, defined as the ratio of transconductance ($\partial I_D/\partial V_{GS}$) to drain current (i.e., g_m/I_D), is used to achieve the desired value of transconductance (speed) by effectively utilizing the current (power). Figure 6 illustrates the TGF of all the devices under consideration. It is clearly shown in Fig. 6 that the device efficiency of D3/D4 is considerably higher than that of D1/D2. It is evaluated that the device efficiency of D3/D4 is 2.31 times higher than D1/D2 @$I_{ds} = 10^{-8}$ A. This enhanced device efficiency of D3/D4 is offered by the improved g_m and I_{ds} by integration of SP and GME engineering.

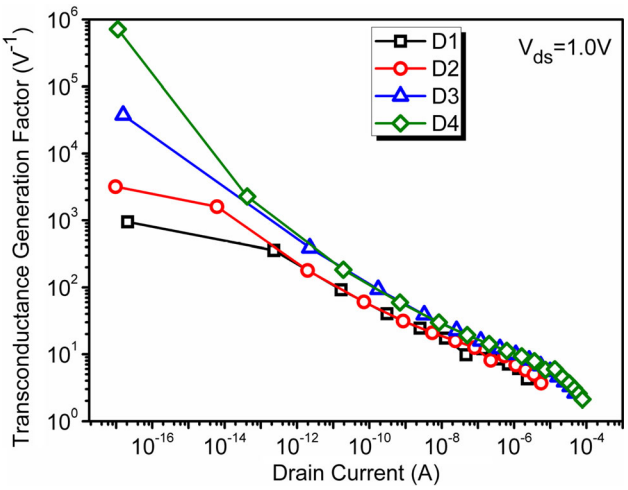


Fig. 6 Transconductance generation factor of all the devices i.e., D1, D2, D3 and D4 at V_{ds} = 1.0 V

3.2 Influence of interface fixed charges (FC)

From Sect. 3.1, it is examined that the D4, i.e., the device with integrated SP and GME on DG-TFET, delivers the superior analog performance in comparison with its conventional counterparts, thus, in this subsection, the impact of FC on the device, D1 and D4, in terms of analog/RF performance has been investigated.

The BTBT at the source and channel/pocket junction for D1/D4 followed by the drift–diffusion of electrons results in drain current in TFET. Further, the band bending at the source and channel/pocket junction for D1/D4 becomes steeper with gate bias. To analyze the impact of interface FC on transfer characteristics, drain current of D1 and D4 under the presence of different polarity (negative/positive) FC is investigated and is depicted in Fig. 7a. Results clearly show that the presence of FC alters the device characteristics from their ideal expected values, i.e. for N_f= 0 cm^{-2}. It is obtained that the positive/negative FCs increases/decreases the drive current. This variation in drain current with FC is due to the fact that the presence of positive/negative polarity FC reduces/enhances the flatband voltage (V_{fb}) (in accordance with Eq. (1)), and thereby

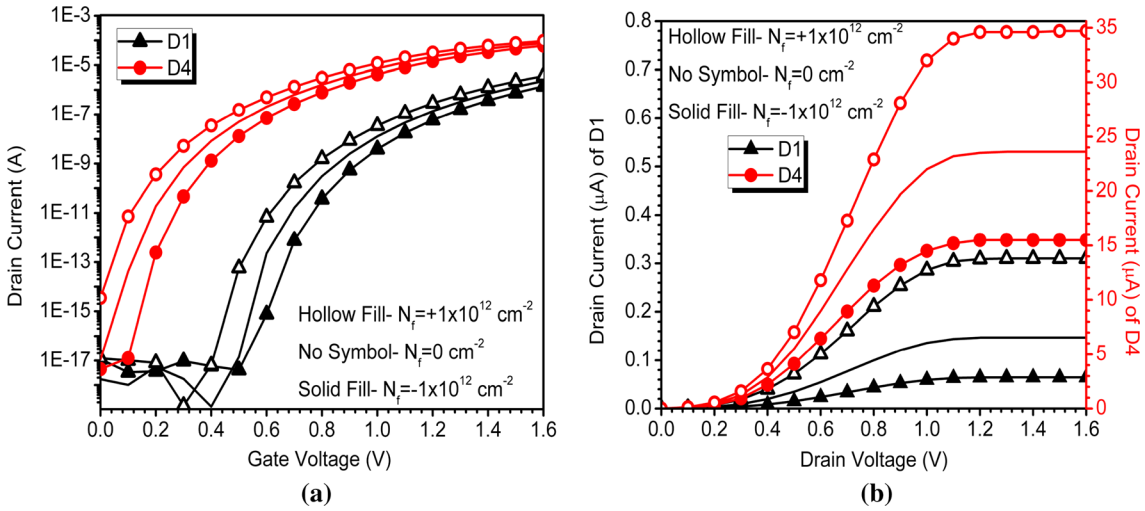


Fig. 7 **a** Transfer characteristics and **b** output characteristics of D1 and D4 depicting the impact of positive, negative and absence of FC

enhances/reduces the effective gate bias (V_{eff}= V_{gs} − V_{fb}) in the channel.

$$\Delta V_{fb} = \frac{qN_f}{C_{ox}} \tag{1}$$

This enhanced/reduced V_{eff} at the tunneling junction enhances/reduces the drain current in DG-TFET. Moreover, it is obtained that the subthreshold current is highly impacted with these FCs as compared to the superthreshold regime for both the devices. It is evaluated that positive FC enhances the subthreshold current by many orders i.e., from an order of 10^{-17} A to 10^{-15} A. This considerably high increase in subthreshold drain current due to the positive FC degrades the device performance significantly. Further, it is clearly shown in Fig. 7a that the variations in drain current due to the positive/negative FC is appreciably reduced in D4 as compared to D1. This improved reliability in the superthreshold regime of D4 is attributed to the integration of GME and SP. The GME utilizes the lower work function metal near the source side and thus have a lower flatband voltage (V_{fb}). This reduced V_{fb} with GME improves the device reliability of D1. Furthermore, the variations in output characteristics with FC for both the devices is shown in Fig. 7b. Results shows that the drain current decreases/increases for negative/positive FC. Its main reason is due to the positive FC reduces the V_{fb} and thereby enhances the effective gate bias in the channel. Further, it is evaluated for D1 that the drain current increases/decrease by 2.10/2.28 times with positive/negative FC. Additionally, the increase/decrease is 1.46/1.52 times for D4, this reduced variations in drain current is offered by the GME and SP engineering. Thus, the reduced variation in drain current for D4 signifies D4 as more reliable device against FC. Thus, along with improved analog/RF performance, D4 offers better reliability. The impact of the presence of interface FC on the threshold voltage (V_{th}), subthreshold swing (SS) and current switching ratio on D1 and D4 is tabulated in Table 2. As already discussed, that the V_{th} of D1 is considerably reduced with the integration of SP and GME, i.e., for D4. Further, the positive/negative FC reduces/enhances the V_{th} for both D1 and D4. This decrease/increase in V_{th} is attributed to the corresponding enhancement/reduction of drain current as described in Fig. 7a.

Further, it is analyzed that the positive FC degrades the SS significantly, owing to the increasing OFF-state current of the device as discussed in Fig. 7a. Further, the current switching ratio, i.e., evaluated as I_{ON}/I_{OFF}, increases/decreases for negative/positive IFC in the case of D4. For D4, the positive FC reduces the I_{ON}/I_{OFF} ratio by many orders owing to the considerably high degradation in the subthreshold region. However, the I_{OFF} in D4 is marginally reduced with negative FC that is translated in marginally enhancement in the I_{ON}/I_{OFF} ratio.

Figure 8a apparently depicts the variations in transconductance (g_m) due to positive/negative FC in comparison with the case where there are no trap charges, for both the devices D1 and D4. It is evident from the plot that the g_m of D1 is considerably small than D4, that is essentially the reason that g_m of D1 is multiplied by 10 so as to display on the same scale. Further, the g_m of both D1 and D4 increases/decreases with the presence of positive/negative FC. This variation in g_m is similar to the drain current, as the g_m is the first order derivative of the drain current. Figure 8b depicts the variations in output conductance due to the presence of FC of different polarity on both D1 and D4. The higher drain current of output characteristics, in turn, makes the output conductance g_d higher for D4 in comparison with D1. However, the analog circuit designers demand low g_d device, to account for the higher intrinsic

Table 2 Variations in V_{th}, SS, and I_{ON}/I_{OFF} of D1, and D4 due to FC

Interface FC	Negative FC $N_f = -1 \times 10^{12}$ cm^{-2}	No FC $N_f = 0$ cm^{-2}	Positive FC $N_f = +1 \times 10^{12}$ cm^{-2}
V_{th} (V)			
D1	1.24	1.16	1.08
D4	0.61	0.53	0.45
SS (mV/decade)			
D1	33.37	14.57	25.32
D4	13.43	27.96	30.14
I_{ON}/I_{OFF}			
D1	5.27×10^9	7.79×10^{10}	2.32×10^{10}
D4	3.23×10^{12}	1.97×10^{12}	9.16×10^9

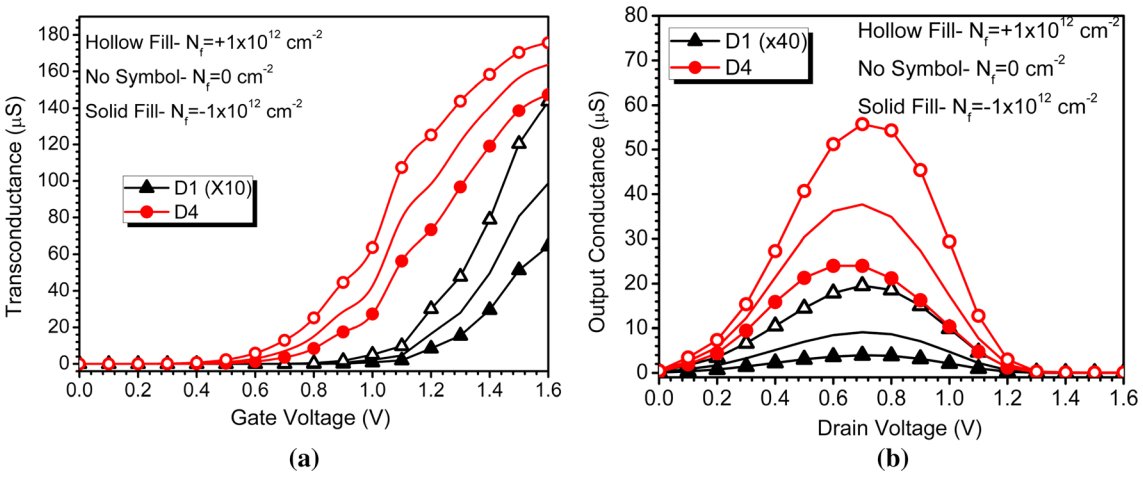


Fig. 8 **a** Transconductance and **b** output conductance of D1 and D4 depicting the impact of positive, negative and absence of FC

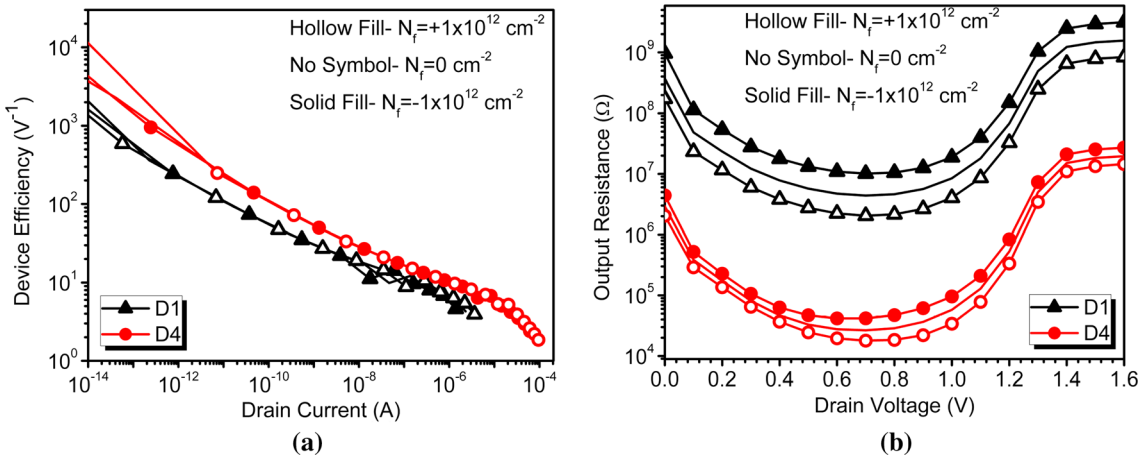


Fig. 9 **a** Device efficiency and **b** output resistance of D1 and D4, depicting the impact of positive, negative and absence of FC

gain A_v. Further, as described beforehand for output characteristics the g_d of both the devices increases/decreases for positive/negative FC.

Figure 9a shows the variations in device efficiency (i.e., the ratio of g_m and I_{ds}), also known as transconductance generation factor (TGF), of D1 and D4 due to the presence of FC. From the plot shown in Fig. 9a it is evident that at lower values of drain current, i.e. in the subthreshold region, TGF for D4 is high as compared to D1. This is because TGF has inverse reliance with the drain current

and in the sub-threshold region drain current is significantly small. So, initially, TGF is having its peak value and with an increase in drain current its value starts decreasing. Also, due to the increase/decrease in g_m and I_{ds} with positive/negative FC, discussed beforehand, device efficiency of both the devices increases/decreases with positive/negative FC. It must be noted that the impact of FC is only marginally on device efficiency. Apparently, from the plot, it is evident that TGF of both D1 and D4 TFET is very high of the order of $\approx 10^4$ at the sub-threshold region, that is much higher as compared to MOSFET, with TGF limited to 40 V^{-1}. The plot in Fig. 9b, shows the impact of FC on the drain resistance or output resistance (R_{out}) as a function of drain voltage. Output resistance is inverse of the output conductance ($g_d = \partial I_D/\partial V_{DS}$). It is apparent from the plot that R_{out} of D4 is smaller in order as compared to D1. This is due to the fact that the conduction is higher in D4 as compared to D1 and output resistance is inverse of g_d. Additionally, the intrinsic gain ($A_v = g_m/g_d$) directly depends on the transconductance which is directly related to TGF. So, higher value of TGF will compensate for lower value of output resistance in order to achieve high intrinsic gain. Further, the inverse reliance of g_d with R_{out} is responsible for the increase/decrease of R_{out} under the presence of negative/positive FC.

The impact of FC on the parasitic gate capacitances as a function of gate voltage is depicted in the plot shown in Fig. 10a, b. It is apparent from the results that the gate-source capacitance increase/decreases only slightly for negative/positive FC. Further, the parasitic capacitances of D4 are much higher than D1. This higher parasitic capacitance of D4 is due to the heavily doped SP. The low parasitic capacitance of any device is desirable as the parasitic capacitance has an inverse relation with the device speed and also with power dissipation (Mookerjea et al. 2009). Thus, the marginal variation in parasitic capacitances with FC signifies the immunity of devices against FC, and thereby the indicates their reliability. Additionally, the plot indicates that initially gate-source capacitance (C_{gs}) is dominant up to the threshold voltage and then gate-drain capacitance (C_{gd}) takes the lead. This is because the formation of inversion layer will appear at the drain first and then it is extended towards the source region, which results in the reduction of coupling between the gate and source (Gnani et al. 2015). So, with increase in V_{gs}, the gate-source capacitance decreases. Further, the total gate capacitance (C_{gg}) is the sum of the C_{gd} and the C_{gs}. From the plot shown in Fig. 10b it is evident that the attributes of total gate parasitic capacitance as a function of gate voltage is similar to the gate-drain capacitance. The similar plot of C_{gd} and C_{gg} is attributed to the dominant contribution of C_{gd} in case of TFET in comparison with C_{gs} (Mookerjea et al. 2010). This is clearly inferred from Fig. 10b that the parasitic capacitance changes marginally under the presence of FC, thereby signifies the device reliability.

The plot shown in Fig. 11a, b depicts the impact of the FC on the parasitic capacitances as a function of drain voltage. It is evident from the plots that the gate parasitic capacitances increase/decrease only minuscule with the positive/negative FC. For any device, these parasitic capacitances should be as minimum as possible because they lead to more power dissipation and also affects the speed of operation. Apparently, it is evident from the plot that the parasitic capacitance increases in D4 as compared to D1 due to the formation of channel/pocket junction, that add on to the total gate capacitance. At the lower values of drain voltage, C_{gs} is significantly small and increases with increase in the drain voltage. This is because for higher value of drain voltage the pinch-off of the channel occurs towards the source region. Vice versa is for C_{gd}, i.e.;

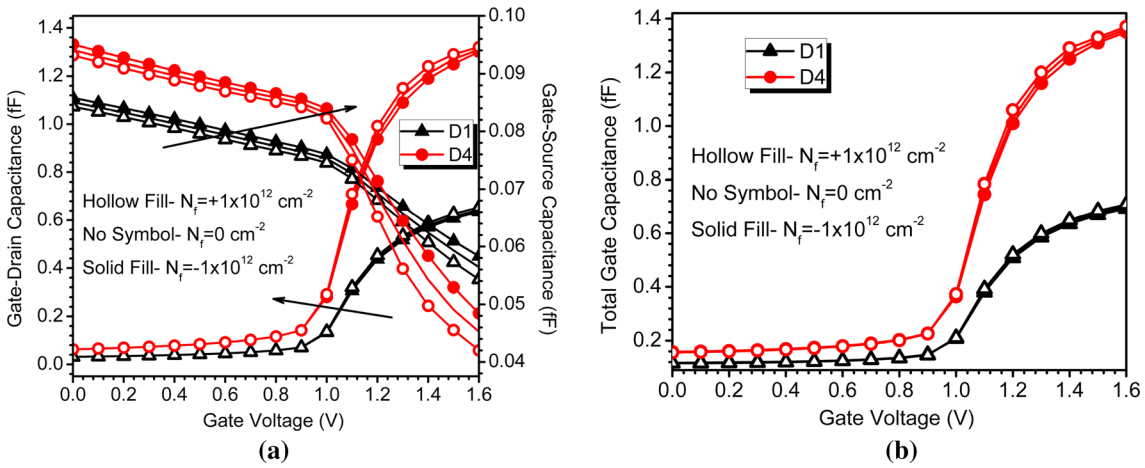


Fig. 10 **a** Gate-drain capacitance, gate-source capacitance and **b** total gate capacitance of D1 and D4 as a function of gate bias, depicting the impact of positive, negative and absence of FC

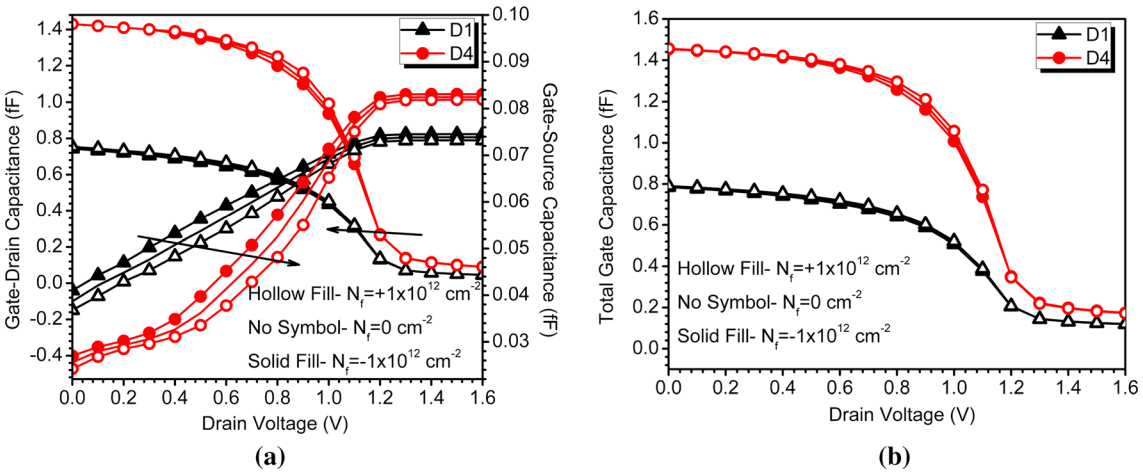


Fig. 11 **a** Gate-drain capacitance, gate-source capacitance and **b** total gate capacitance of D1 and D4 as a function of drain bias, depicting the impact of positive, negative and absence of FC

initially it has high value, and with increase in drain voltage it decreases (Mookerjea et al. 2009). This is due to the fact that with increasing drain voltage the coupling towards the drain also diminishes. Further, the total gate capacitance C_{gg} is analyzed for both the devices, under the presence of different FC and is shown in Fig. 11b. It is evident that the attributes of total gate parasitic capacitance as a function of drain voltage is similar to the gate-drain capacitance, owing to the considerably small C_{gs}. Further, the impact of FC is negligibly small on the parasitic capacitances showing the immunity of device against the FC.

For high-speed RF applications, the cut off frequency (f_T) of device is the most desired design parameter which is directly related to the transconductance and inversely to the parasitic capacitances. f_T of the device is an important parameter to analyze the RF performance of any device and is evaluated by (Marjani and Hosseini 2015)

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \tag{2}$$

It is inferred from Eq. (2) that the higher parasitic capacitances may degrade the f_T of the device. In our study, it is realized that the integration of SP and GME engineering degrades the C_{gg} as discussed in Figs. 10 and 11, however, it also enhances the drain current considerably and thereby the g_m (depicted in Fig. 8a). The plot of f_T evidently shows that the degradation in C_{gg} (of D4 as compared to D1) is overpowered by the enhanced drain current (of D4 as compared to D1). This results in the considerable enhancement in f_T of D4 as compared to D1. Further, the impact of FC on f_T is also depicted in Fig. 12a that reveals that for both the devices the f_T increases/decreases due to the presence of positive/negative FC. This variation in f_T is similar to the variation in g_m, as enhancement in g_m overpowers the variations in C_{gg}. It is evaluated that the increase/decrease in f_T is 79.89%/48.02% (in D1) and only 23.85%/23.89% (in D4) with positive/negative FC at V_{gs} = 1.2 V. The reduced variations in f_T in D4 implies the better reliability of D4 in comparison with D1.

Another important parameter to inspect the device RF performance is the transconductance frequency product (TFP) i.e. primarily the product of f_T and the device efficiency (g_m/I_{ds}). The variations in TFP for both the devices under presence of FC of different polarity is depicted in Fig. 12b. It is clearly shown in the Fig. 12b that the TFP of D4 is appreciably higher than D1 by many orders, attributed to the enhanced f_T and device efficiency of D4. This higher TFP possessed by D4 can be beneficial for moderate to high-speed device designs and is simultaneously good to maintain the linearity of the device (Pradhan et al. 2014). Moreover, the impact of FC on TFP is the communal effect of both device efficiency and the f_T. Thus, the TFP increases/decreases with the presence of positive/negative FC.

4 Conclusions and future work

In this work, the integrated merits of gate metal engineering and source pocket have been incorporated on DG-TFET to obtain GME-SP-DG-TFET. To insight the device physics the electrical/analog performance of GME-SP-DG-TFET along with its conventional counterparts, i.e., GME-DG-TFET, SP-DG-TFET, and DG-TFET have been examined by calibrated SILVACO TCAD setup. The investigation reveals that the collective merits of SP and GME outperforms its conventional counterparts and simultaneously considerably overcome the major

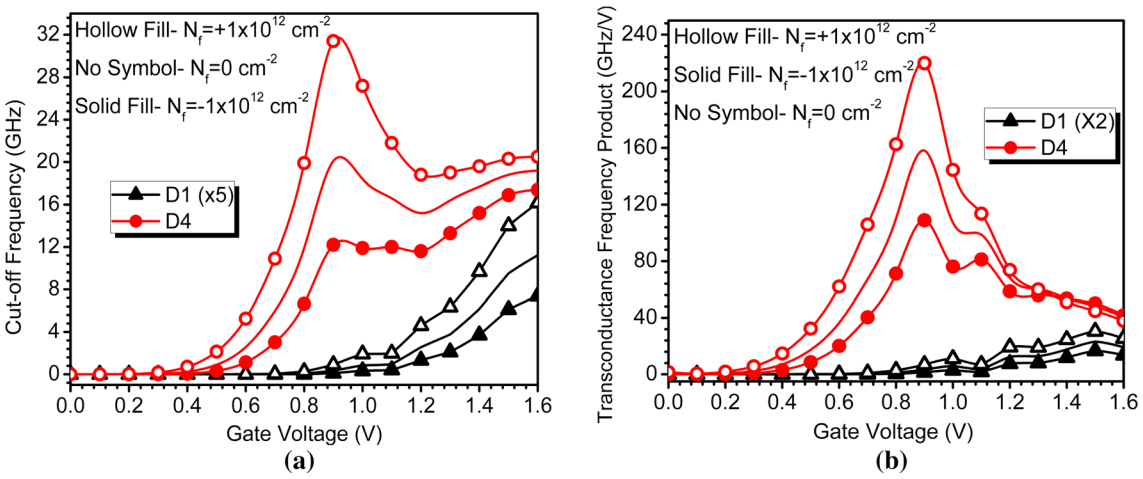


Fig. 12 **a** Cut-off frequency and **b** transconductance frequency product of D1 and D4 as a function of drain bias, depicting the impact of positive, negative and absence of FC

roadblock of TFET. It is evaluated that the I_{ON}/I_{OFF} ratio improves from an order of 10^{10} to 10^{12} for GME-SP-DG-TFET that makes it an appropriate candidate for high switching applications. Subsequently, the reliability of the GME-SP-DG-TFET is examined and is simultaneously compared with the conventional DG-TFET. In this respect, the device analog/RF performance is inspected in the presence of FC. It is revealed that the presence of positive FC enhances the subthreshold current considerably by many orders and thus are more noxious than the negative FC. Further, it is obtained that the FC dominantly alters the device characteristics in the subthreshold regime as compared to the superthreshold regime. It is assessed in the ON state, that the increase/decrease in f_T is 79.89%/48.02% (in D1) and only 23.85%/23.89% (in D4) with positive/negative FC. The significant reduction in variations in f_T in GME-SP-DG-TFET implies its better reliability. Thereby, the study reveals better reliability of the GME-SP-DG-TFET along with its improved analog/RF performance. Thus, the lower I_{OFF} (10^{-17} A), higher I_{ON}/I_{OFF} ratio (10^{12}), lower SS (27 mV/decade), lower V_{th} (0.62 V) of GME-SP-DG-TFET will effectively meet the need of low power and high switching speed applications. This detailed analysis would be beneficial for designing and optimizing a reliable TFET for low power analog/RF applications.

In the future, one can utilize the proposed GME-SP-DG-TFET as a label-free biosensor, by etching the oxide layer that will act as the binding site for the biomolecules. The biomolecules will bind in the oxide cavity and will thereby alter the device characteristics. The altered device characteristics under the presence of biomolecules in comparison with the case when there is no biomolecule present may be calibrated for detection of biomolecules.

Acknowledgements Authors would like to thank Microelectronics Research Lab, Department of Applied Physics, Delhi Technological University and all members of VLSI Center of Excellence at Chitkara University for their support in this work.

References

Biswal SM, Baral B, De D, Sarkar A (2019) Simulation and comparative study on analog/RF and linearity performance of III–V semiconductor-based staggered heterojunction and InAs nanowire (nw) Tunnel FET. Microsyst Technol 25(5):1855–1861

Boucart K, Ionescu AM (2007) Length scaling of the double gate tunnel FET with a high-k gate dielectric. Solid-State Electron 51:1500–1507

Cao W, Yao C, Jiao G, Huang D, Yu H, Li M-F (2011) Improvement in reliability of tunneling field-effect transistor with pnin structure. IEEE Trans Electron Dev 58:2122–2126

Choi WY, Park B-G, Lee JD, Liu T-JK (2007) Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec. IEEE Electron Dev Lett 28:743–745

Cui N, Liang R, Xu J (2011) Heteromaterial gate tunnel field effect transistor with lateral energy band profile modulation. Appl Phys Lett 98:142105

Datta E, Chattopadhyay A, Mallik A, Omura Y (2020) Temperature dependence of analog performance, linearity, and harmonic distortion for a ge-source tunnel FET. IEEE Trans Electron Dev 67:810–815

Gnani E, Baravelli E, Gnudi A, Reggiani S, Baccarani G (2015) Capacitance estimation for InAs Tunnel FETs by means of full-quantum k· p simulation. Solid State Electron 108:104–109

Guillaumot B et al (2002) 75 nm damascene metal gate and high-k integration for advanced CMOS devices. In: Digest. International electron devices meeting, 2002. IEEE, pp 355–358

Jiao G et al (2009) New degradation mechanisms and reliability performance in tunneling field effect transistors. In: 2009 IEEE international electron devices meeting (IEDM), 2009. IEEE, pp 1–4

Joshi T, Singh Y, Singh B (2020) Extended-source double-gate tunnel FET with improved DC and analog/RF performance. IEEE Trans Electron Dev. https://doi.org/10.1109/ted.2020.2973353

Kim JH, Kim S, Park B-G (2019) Double-gate TFET with vertical channel sandwiched by lightly doped Si. IEEE Trans Electron Dev 66:1656–1661

Kumar N, Raman A (2019) Low voltage charge-plasma based dopingless tunnel field effect transistor: analysis and optimization. Microsyst Technol. https://doi.org/10.1007/s00542-019-04666-y

Lho YH, Kim KY (2005) Radiation effects on the power MOSFET for space applications. ETRI J 27:449–452

Li W, Woo JC (2020) Vertical P-TFET with a P-type SiGe pocket. IEEE Trans Electron Dev 67(4):1480–1484

Liu J, Wen H, Lu J, Kwong D-L (2005) Dual-work-function metal gates by full silicidation of Poly-Si with Co–Ni bi-layers. IEEE Electron Dev Lett 26:228–230

Long W, Ou H, Kuo J-M, Chin KK (1999) Dual-material gate (DMG) field effect transistor. IEEE Trans Electron Dev 46:865–870

Lu B, Lu H, Zhang Y, Zhang Y, Cui X, Lv Z, Liu C (2018) Fully analytical carrier-based charge and capacitance model for hetero-gate-dielectric tunneling field-effect transistors. IEEE Trans Electron Dev 65:3555–3561

Madan J, Chaujar R (2016) Interfacial charge analysis of heterogeneous gate dielectric-gate all around-tunnel FET for improved device reliability. IEEE Trans Dev Mater Reliab 16:227–234

Madan J, Chaujar R (2017a) Gate drain underlapped-PNIN-GAA-TFET for comprehensively upgraded analog/RF performance. Superlattices Microstruct 102:17–26. https://doi.org/10.1016/j.spmi.2016.12.034

Madan J, Chaujar R (2017b) numerical simulation of N^+ source pocket PIN-GAA-tunnel FET: impact of interface trap charges and temperature. IEEE Trans Electron Dev 64:1482–1488

Madan J, Chaujar R (2018) Temperature associated reliability issues of heterogeneous gate dielectric—gate all around—tunnel FET. IEEE Trans Nanotechnol 17:41–48

Madan J, Gupta RS, Chaujar R (2017a) Mathematical modeling insight of hetero gate dielectric-dual material gate-GAA-tunnel FET for VLSI/analog applications. Microsyst Technol 23:4091–4098. https://doi.org/10.1007/s00542-016-2872-9

Madan J, Gupta RS, Chaujar R (2017b) Performance investigation of heterogeneous gate dielectric-gate metal engineered–gate all around-tunnel FET for RF applications. Microsyst Technol 23:4081–4090. https://doi.org/10.1007/s00542-016-3143-5

Madan J, Shekhar S, Chaujar R (2017c) Gate metal engineered heterojunction DG-TFETs for superior analog performance and enhanced device reliability. In: 2017 conference on information and communication technology (CICT). IEEE, pp 1–4

Mallik A, Chattopadhyay A (2012) Tunnel field-effect transistors for analog/mixed-signal system-on-chip applications. IEEE Trans Electron Dev 59:888–894

Marjani S, Hosseini SE (2015) Radio-frequency small-signal model of hetero-gate-dielectric pnpn tunneling field-effect transistor including charge conservation capacitance and substrate parameters. J Appl Phys 118:095708

Mookerjea S, Krishnan R, Datta S, Narayanan V (2009) On enhanced Miller capacitance effect in interband tunnel transistors. IEEE Electron Dev Lett 30:1102–1104

Mookerjea S, Mohata D, Mayer T, Narayanan V, Datta S (2010) Temperature-dependent I-V characteristics of a vertical In0.53Ga0.47As tunnel FET. IEEE Electron Dev Lett 31:564–566

Nagavarapu V, Jhaveri R, Woo JC (2008) The tunnel source (PNPN) n-MOSFET: a novel high performance transistor. IEEE Trans Electron Dev 55:1013–1019

Nirschl T et al (2006) Scaling properties of the tunneling field effect transistor (TFET): device and circuit. Solid State Electron 50:44–51

Pala M, Esseni D, Conzatti F (2012) Impact of interface traps on the IV curves of InAs tunnel-FETs and MOSFETs: a full quantum study. In: 2012 IEEE international electron devices meeting (IEDM), 2012. IEEE, pp 6.6.1–6.6.4

Poindexter EH (1989) MOS interface states: overview and physicochemical perspective. Semicond Sci Technol 4:961

Polishchuk I, Ranade P, King T-J, Hu C (2001) Dual work function metal gate CMOS technology using metal interdiffusion. IEEE Electron Dev Lett 22:444–446

Pradhan K, Mohapatra S, Sahu P, Behera D (2014) Impact of high-k gate dielectric on analog and RF performance of nanoscale DG-MOSFET. Microelectron J 45:144–151

Qiu Y, Wang R, Huang Q, Huang R (2014) A comparative study on the impacts of interface traps on tunneling FET and MOSFET. IEEE Trans Electron Dev 61:1284–1291

Sant S, Moselund K, Cutaia D, Schmid H, Borg M, Riel H, Schenk A (2016) Lateral InAs/Si p-type tunnel FETs integrated on Si—part 2: simulation study of the impact of interface traps. IEEE Trans Electron Dev 63:4240–4247

Saurabh S, Kumar MJ (2011) Novel attributes of a dual material gate nanoscale tunnel field-effect transistor. IEEE Trans Electron Dev 58:404–410

Shekhar S, Madan J, Chaujar R (2018) Source/gate material-engineered double gate TFET for improved RF and linearity performance: a numerical simulation. Appl Phys A 124:739

SILVACO I (2011) ATLAS User's Manual Santa Clara, CA, Ver 5

Singh PK, Baral K, Kumar S, Chander S, Tripathy MR, Singh AK, Jit S (2020) Source pocket engineered underlap stacked-oxide cylindrical gate tunnel FETs with improved performance: design and analysis. Appl Phys A 126:166

Trabzon L, Awadelkarim O (1998) Damage to n-MOSFETs from electrical stress relationship to processing damage and impact on device reliability. Microelectron Reliab 38:651–657

Tura A, Zhang Z, Liu P, Xie Y-H, Woo JC (2011) Vertical silicon pnpn tunnel nMOSFET with MBE-grown tunneling junction. IEEE Trans Electron Dev 58:1907–1913

Villalon A, Le Carval G, Martinie S, Le Royer C, Jaud M-A, Cristoloveanu S (2014) Further insights in TFET operation. IEEE Trans Electron Dev 61:2893–2898

Wang P-F et al (2004) Complementary tunneling transistor for low power application. Solid State Electron 48:2281–2286

Wei S, Zhang G, Geng L, Shao Z, Yang C-F (2018) Comparison of the performance improvement for the two novel SOI-tunnel FETs with the lateral dual-gate and triple-gate. Microsyst Technol. https://doi.org/10.1007/s00542-018-4018-8

Zhang Z et al (2005) Integration of dual metal gate CMOS on high-k dielectrics utilizing a metal wet etch process. Electrochem Solid State Lett 8:G271–G274

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)

Language Detection of Text Document Image

Dr. N. Jayanthi
Asst. Prof., ECE
Delhi Technological University
New Delhi, India
njayanthi@dce.ac.in

Harsha
B. Tech, ECE
Delhi Technological University
New Delhi, India
harsha15soni@gmail.com

Naman Jain
B. Tech, ECE
Delhi Technological University
New Delhi, India
naman_bt2k15@dtu.ac.in

Ishpreet Singh Dhingra
B. Tech, ECE
Delhi Technological University
New Delhi, India
ishpreet.dhingra@gmail.com

***Abstract* — Language detection is an important feature for developing a Multilingual OCR. Lots of research has been done on Languages such as English, French, Spanish but very limited work has been performed on Indian languages such as Hindi, Tamil etc. due to their complex nature. In this paper we are presenting a method to detect the language of a text document image. The language chosen are such that they are of different scripts. The process of language detection involves segmentation of individual characters from the document and then labelling them to a particular language. A Convolutional Neural Network is trained on characters of different languages. A text document is classified on the basis of what language its individual characters belong to. This is achieved by first segmenting individual characters of a document and then classifying each of them separately.**

Keywords - Convolutional Neural Network, Pre-processing, Segmentation, Shirorekha, Classification, Training

I. INTRODUCTION

Language detection is the process of classifying a document on the basis of its language. Identifying the language of a document is a key step in a document reading system. The human visual system is one of the wonders of the nature. Most people effortlessly recognize language written in a document if they are aware of that language but they are unable to do this for an unknown language. To automate this process, we attempt to build a model which can recognize the language of a text document.

Most of the languages have separate OCR models. Therefore, we require to first recognize the language and then classify its characters, to convert a text document into its digital copy. In this paper, we are trying to solve this problem at a basic level by classifying a text document as one of India's 3 major languages (with number of speakers) – English (12.9 crores), Hindi (69.2 crores) and Tamil (7.7 crores) [1].

For this purpose, like humans, we first make the system learn characters of different languages using convolutional neural networks. The idea is to take a large number of characters of each language as training data and then develop a system which can learn from those training examples. But testing document doesn't contain individual characters rather, it contains a set of characters. So, in order to test the document, we first extract individual characters from the document and then detect the language of the document on the basis of those extracted character using our trained model.

The choice of languages was done on the basis of their scripts. Our model will only be able to detect languages of different scripts. Therefore, we have chosen English (Latin script), Hindi (Devanagari script) and Tamil (Abugida script).

A.Motivation

We are moving towards a generation of machine learning and artificial intelligence. A generation in which computers will be capable of doing every such thing which humans can do. Human visual system is one of the wonders of world and we are trying to mimic it. In order to start performing any activity a system first needs to know what it is supposed to do. Humans have ears and eyes to receive information and a brain to act accordingly. Similarly, for computers, recognition of words on its own would act as its eyes and voice recognition as ears. Understanding language is one of the crucial steps for doing any activity. Even in case of humans we first understand the language and then process the task. Computers in a similar way needs to learn the language and then once learning is completed it should be able of recognizing the language. So, as we now know how important it has become for machines to recognize and understand human language we decided to work in this field.

B.Proposed Methodology

978-1-7281-5475-6/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

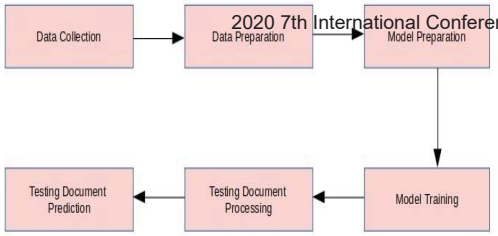


Fig. 1 Methodology

a) *Data Collection* - Images of all characters of English, Hindi, Tamil are taken as training data. 13000 images each of English (62 characters which consists of 26 capital and 26 small letters and 10 numeric digits.) [2], Hindi (47 characters) [3], Tamil (124 characters) [4] are taken.

Some examples of characters of each language:



Fig. 2 English dataset



Fig. 3 Hindi dataset



Fig. 4 Tamil dataset

b) *Data Preparation* - As the data for all 3 languages are obtained from different sources, they are not uniform. So, some pre-processing like thresholding, blank space removal, resizing etc. is done to make the data uniform. Each of the training image is resized to 25x25x1.

c) *Model* - A CNN with 4 convolutional layers and 3 Max pooling layers is formed. ReLu and Sigmoid are used as activation functions. Adam is the optimizer used. The model is trained on individual characters of each language. The label for every character is the language it belongs to. So, our model has 3 output nodes which give us the probability values for the input character for all of the 3 languages.

d) *Testing Document Processing* - Thresholding, Gaussian Blur, *Clahe* transform are applied on the testing image. Since we have used individual characters for training our model, we need to break down our testing image into individual characters for prediction. We have applied contour detection on our image to segment it into different characters. Some further steps are required for Hindi documents because in Hindi document, characters are joined to each other by a line known as *shirorekha*.

e) *Prediction* - After processing the testing document into individual characters, each of the characters is given to the model as input and we get the probability values of it belonging to all 3 languages. The language with the Highest probability value is assigned to it. The count of characters of each language is calculated. The language having the highest number of characters in the document is given as the output.

II. PREPROCESSING

A. Binarization

Grayscale image has pixel values ranging from 0 to 255. This results in large number of feature values which can increase the complexity of our model. Therefore, we require to binarize our image pixel values. For binarization we require a threshold value, such that all the value below it, are changed to 0 and vice-versa for higher values. In this case we have used Otsu thresholding for binarization as it gave us the best results compared to other techniques [5].

In Otsu thresholding, a histogram of pixel values is used to find two peaks and a value between these peaks is chosen as the threshold value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

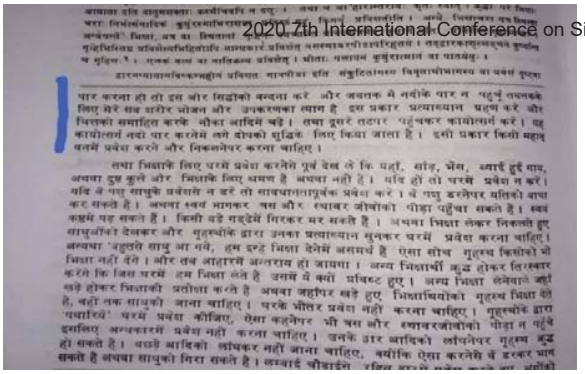


Fig. 5 Original Image

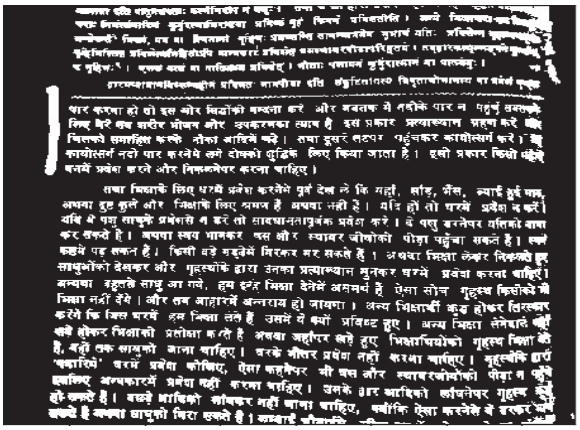


Fig. 6 After thresholding

B. Image processing

Pre-processing is done in order to obtain dataset in a form which has minimum noise, pixel values not limited in one region and minimum skew. Our classifier model is trained to recognize characters; therefore, we need to obtain characters from the image for classification. In order to fetch characters, we use contour detection which finds the characters present in the image document.

a) Processing training image: For improvement in contrast we use histogram equalization of image which is done using *clahe* technique [6]. *Clahe* (Contour limited adaptive histogram equalization) enhances only the key areas (edges, corners etc.).

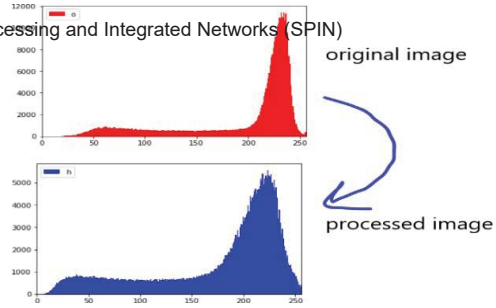


Fig. 7 Histogram before and after clahe

Besides poor contrast, blank spaces also existed in the dataset which resulted into unwanted features. In order to remove these unwanted pixels, we removed blank spaces from our training image resulting into an image with useful features only.

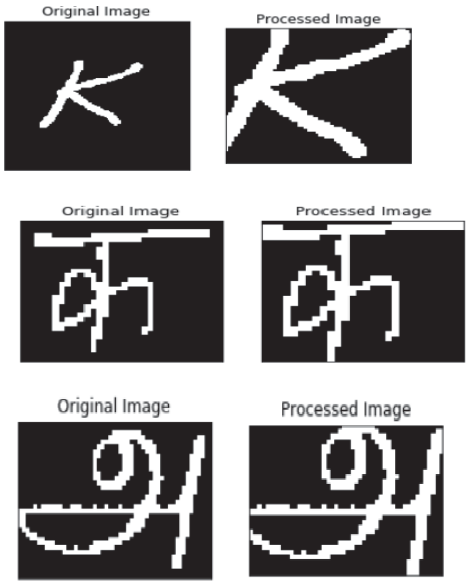


Fig. 8 Characters before and after space removal

b) *Resizing* of the training dataset is also done as during training of model, input images need to be of the same size. We resized the training images to 25 x 25.

c) *Processing testing image:* Besides noise and poor contrast testing image may also be skewed [7]. Therefore, we apply thresholding and histogram equalization to the testing image as well. In order to remove *skew* we use one the various available skew detection and correction tools.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

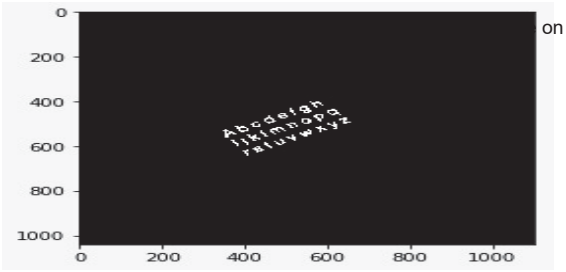


Fig. 9 Before skew removal

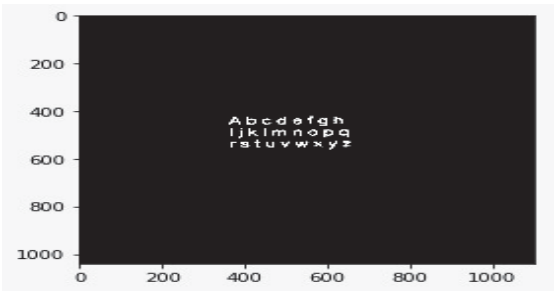


Fig. 10 After skew removal

Since our model only recognizes characters. So, we need to extract characters from the testing image. In order to extract characters, we apply contour detection. Contour detection returns all the characters which are not connected, so it applies a constraint to our work that we cannot apply contour detection on images which contain cursive writing. This constraint even extends to Hindi documents due to presence of *shirorekha* on each word. But we have handled this problem (explained later) in case of Hindi document by first removing the *shirorekha* and then finding the contour.

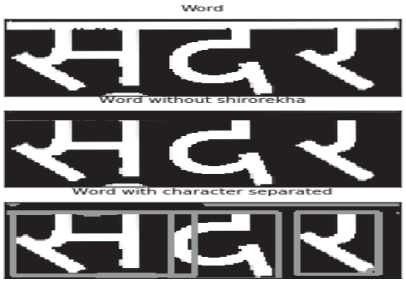




Fig. 11 Images before and after contour detection

d) *Segmenting Individual character:* In case of Tamil and English document we can easily detect individual characters by applying contour detection once. In case of Hindi we get words instead of individual characters. So, to extract individual characters we first locate the *shirorekha*. This is done by finding the row which has maximum number of white pixels in the word which is obtained after contour detection. After locating *shirorekha,* we remove it by converting all pixels containing the *shirorekha* into black pixels. This leaves us with two parts, above and below the *shirorekha*. We consider only the lower part for classification. Then we apply contour detection for the second time to get contours around individual characters [8].

Since the length of contour around a word is much greater than its width, the above method is applied only when the length to width ratio of the contour is higher than a certain threshold which is only in the case of Hindi. Threshold value was set as a hyper parameter and a value of 1.75 gave us the best results.



Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

Fig. 12 Contour detection on word after removing shirorekha

After characters are extracted from the testing image, we resize them to 25 x 25. After this we found that characters obtained were thick, that is they consisted of more white pixels than required so, in order to get a sharper character, we needed to perform *morphological transformation.* Basic morphological operator used to decrease thickness is *Erosion*.

Erosion basically erodes away the boundaries of the foreground object. So, the thickness or size of the foreground object decreases or simply white region decreases in the image [9].

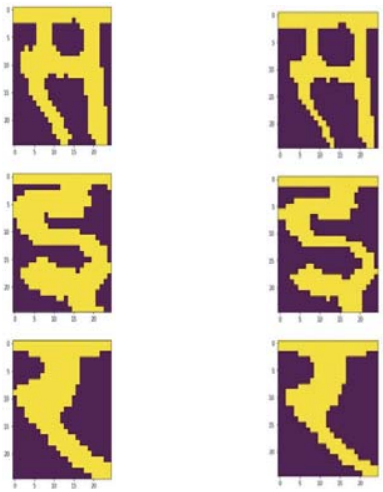


Fig. 13 Characters before and after erosion

III. CLASSIFICATION

A. Filtering noisy contours

With the text, some noisy elements are also present in the image. The noise in the image is mainly contributed by improper binarization of images which is generally because of varying luminous intensity on the document, abrupt pen strokes in a handwritten document etc. So, when contour detection is applied, these noisy elements also detected as separate contours, thereby inducing error during the prediction.

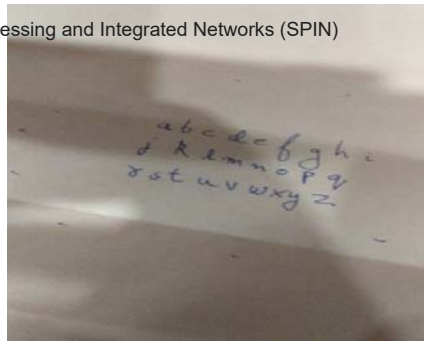


Fig. 14 Noisy image

The noise due to variation in luminosity is eliminated by using proper thresholding and blurring techniques. For other noisy elements, we observed that their size is very small as compared to the actual characters written in the document. Thus, after the contours are identified in the input document, we can filter these noisy elements on the basis of a threshold area value of the identified contours.

A threshold value was found by observing the sizes of noisy elements in the testing image. It was observed that contours which were smaller than *0.00015 times* the *area of the document* were mainly due to the noisy elements and therefore were not considered for classification of the document.

B. Model selection

For preparing the model, we have used *Convolutional Neural Network*. A Convolutional Neural Network learns to recognize patterns across space. It learns to recognize features of the image (e.g., lines, curves, edges, corners, etc.) and then it learns to combine these features to recognize larger structures (e.g., faces, objects, etc.). Each input image is passed through a CNN network which contains convolution layers, Pooling layers, etc. to classify an object.

A kernel is basically a matrix with which the input image is convoluted. The values of the kernel matrix are twitched
using a backpropagation algorithm such as gradient descent. Gradient Descent Algorithm tries to minimize the error in classification by updating the kernel matrix values (weights).
Generally multiple kernels are used. Each kernel tries to learn a small level feature of the image. As the layers in the model increases, the size of the input to the layer decreases, the kernels also learn much higher-level features of the image.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

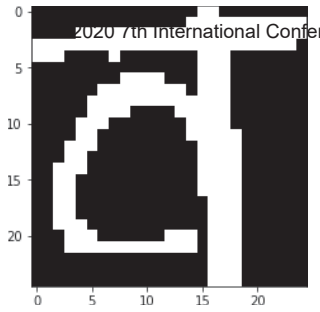


Fig. 15 Sample training image for the model



Fig. 16 Outputs of first layer using 64 kernels

Pooling layer reduces the size of the representation of the features to reduce the number of parameters and computations in the network. It reduces the dimensionality of every map while retaining the important information for further classification [10].

Our CNN Model consists of *four* Convolution layers and *three maxpool* layers with *Relu* and *sigmoid* as the activation functions [11]. First layer has 64 kernels of size 3*3 where the input image has a dimension of 25*25 pixels. Maxpool layer with kernel size of 2*2 is added to progressively decrease the size of the representation. These are followed by subsequent convolution and maxpool layers with continuous reduction in the number of kernels in each convolution layer. At the end *sigmoid* activation function is used to get stochastic outputs for the three output nodes representing the three languages. We used *Adam as* the gradient decent optimizer for the gradient decent algorithm used in backpropagation. Unlike stochastic gradient descent, Adam maintains a different learning rate for each kernel [12].

The pre-processed dataset images of all the three languages were used to train our Convolutional Neural Network Model. The training dataset consisted of *39000* handwritten images of characters from three languages written in varying font weight, angles, coverage. *80%* of characters were used for Training and *20%* for Testing model accuracy of character classification.

After training the prepared model, we obtained a model accuracy of *78%* in 40 learning epochs. The alpha value for Adam optimizer used was 0.001 and the epsilon value was *1e-08*. This trained model was then used to classify the contours detected in the testing document image.

The extracted and processed contour elements were given as inputs to the model for classification. For the testing image, the Model predicts the language (class) of every contour and the counts for the predicted classes are saved. Document is finally classified as the language with the maximum number of predictions.

As the contours were mapped to 25*25 pixels sizes, the width of the characters obtained were thick, we performed Erosion operation. This retains the basic features and reduces the unwanted extra width. This significantly increased the counts of correctly classified contours in the document.

We tested our model on 50 text documents from various sources which includes manual capturing of document image, gathering from internet which exhibited different levels of difficulties for the classifier and observed an *accuracy* of *74%.*

IV. CONCLUSION AND FUTURE WORK

These results illustrate that our model can be applied successfully to detect the language of a text document. There are many factors that affect the performance of the developed language detector. We found that an input image of 25 x 25 gives us the best result. Other types of pre- processing and neural network model may give better result in the future. Character segmentation can be improved to handle documents written in cursive writing [13]. Training on different font type may also improve the performance of our model. In Hindi and Tamil even though the characters are identified correctly, punctuation marks might get classified incorrectly. Also, we are not using *matras* for classification which can be incorporated in future to enhance the performance of our model.

The other future enhancement that can be done is using our model for building Hindi OCR [14], English OCR [15], Tamil OCR [16]. As different languages use different methodologies for building their OCRs, language detection is an important step to achieve it. By training on other languages of different scripts, our model can be extended to detect those languages as well. We can use dic-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

tionary to distinguish between languages of the same script by first applying OCR and then checking to which language the word belongs to.

REFERENCES

[1] https://en.wikipedia.org/wiki/Languages_of_India

[2] https://www.nist.gov/srd/nist-special-database-19

[3] https://www.kaggle.com/jhashanku007/devnagri-hindi-dataset

[4] http://www.jfn.ac.lk/index.php/data-sets-printed- tamil-characters-printed-documents/

[5] Yousefi, Jamileh. (2015). Image Binarization using Otsu Thresholding Algorithm. 10.13140/RG.2.1.4758.9284.

[6] David Ferguson, Phillip & Arslan, Tughrul & Erdogan, Ahmet & Parmley, Andrew. (2008). Evaluation of contrast limited adaptive histogram equalization. (CLAHE) enhancement on a FPGA. 119-122. 10.1109/SOCC.2008.4641492.

[7] Dengel, Andreas & Ahmad, Riaz. (2016). A Novel Skew Detection and Correction Approach for Scanned Documents.

[8] Yadav, Divakar & Sanchez-Cuadrado, Sonia & Morato, Jorge. (2013). Optical Character Recognition for Hindi Language Using a Neural-network Approach. Journal of Information Processing Systems. 9117. 117-140. 10.3745/JIPS.2013.9.1.117

[9] Srisha, Ravi & Khan, Am. (2013). Morphological Operations for Image Processing: Understanding and its Applications.

[10] Albawi, Saad & Abed MOHAMMED, Tareq & ALZAWI, Saad. (2017). Understanding of a Convolutional Neural Network. 10.1109/ICEngTechnol.2017.8308186.

[11] Vehbi Olgac, A & Karlik, Bekir. (2011). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. International Journal of Artificial Intelligence and Expert Systems. 1. 111-122.

[12] Diederik P. Kingma, Jimmy Lei Ba (2015). ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION

[13] Arica, Nafiz & Yarman Vural, Fatos. (2002). Optical character recognition for cursive handwriting. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 24. 801-813. 10.1109/TPAMI.2002.1008386.

[14] Bansal, Veena & M. K. Sinha, R. (2001). A Complete OCR for Printed Hindi Text in Devanagari script. 800-804. 10.1109/ICDAR.2001.953898.

[15] Al-Hassani, Dr. Mustafa. (2013). OPTICAL CHARACTER RECOGNITION (OCR) SYSTEM FOR MULTIFONT ENGLISH TEXTS USING DCT & WAVELET TRANSFORM.

[16] R. Jagadeesh Kannan, R. Prabhakar (2009). A Comparative Study of Optical Character Recognition for Tamil Script.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:06:09 UTC from IEEE Xplore. Restrictions apply.

MTFCT: A task offloading approach for fog computing and cloud computing

Rajni Jindal
Deprtment of CSE
Delhi Technological University
Delhi, India

Neetesh Kumar
ABV-Indian Institute of Information Technology and Management
Gwalior, Madhya Pradesh, India

Hitesh Nirwan
Department of CSE
Delhi Technological University
Delhi, India

***Abstract*— Cloud computing is an important computing paradigm for handling all types of computations, even the smaller ones in the past. But sometimes, it becomes ineffective when the task is to be done in real-time, with very low latency. Therefore, fog computing was introduced as a supplement computing paradigm for cloud computing. Internet of Things based applications perform better with the amalgamation of its and the fog computing. Due to low capacity, when fog can't compute the task on its own, heavy computations are offloaded from fog to cloud. But when to offload the task from fog to cloud is a major decision. The decision is to be made out to offload the tasks from fog to cloud is very crucial, so this paper presents an idea to solve this problem.**

Keywords*— *Cloud Computing; Fog Computing; Task Offloading

I. INTRODUCTION

With the advancement in the technology area, more and more electronically enabled devices, which are capable are connecting with the internet, so the Internet of Things (IoT) devices are increasing day by day. These devices are communicating with each other, so basically, they are generating lots of data [1].

Because of this, the burden on cloud computing to solve computations and to provide the result back in a given time was also increasing day by day.

Cloud computing is an important computing paradigm to handle heavy scientific workflows by providing infinite virtual resources but doesn't work well where latency requirement is low [2].

There may be a case where latency-sensitive applications demand results back in real-time. Cloud may provide an accurate result, but it may return the result late, which may diminish the whole objective.

To solve this issue, fog computing was introduced in 2014 [3]. Fog computing is an online computing paradigm to solve low complexity problems in real-time. It was introduced as a supplement to cloud computing. Its aim was never to replace cloud computing, rather reduce the overload on cloud datacenters and provide fast services to the latency-sensitive application [14].

So it can perform both the operations, storing as well as processing the data near the edge of the network or near the end-user. The interconnection of IoT devices with fog nodes and fog with the cloud is shown in fig 1.



Fig. 1. Three tiers IoT, Fog and Cloud Architecture [11]

Fog computing is decentralized in nature, unlike cloud, which is centralized [4]. It also provides fog computing an upper edge over cloud computing

II. BACKGROUND KNOWLEDGE AND MOTIVATION

According to an estimate, more than 50 billion devices will be connected to the internet by 2020 [5]. The majority of these devices are not able to process their data on their own. So, we need the help of fog computing also along with cloud computing. IoT is the future, but we need to amalgamate it with fog computing. We can not overload cloud datacenters.

Moreover, providing a reliable result is still not an easy task. The performance of fog computing is dependent upon fog

978-1-7281-2791-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:27 UTC from IEEE Xplore. Restrictions apply.

nodes placement as well as the resources provided to the fog nodes. Fog nodes are generally much closer, at the edge of the network [6]. Therefore, their latency time is very less as compared to cloud datacenters.

Fog nodes are small data centers that provide the computational capability to these latency-sensitive devices [7]. They are generally placed between the end devices and the cloud data centers in the network — the closeness of these nodes with end devices benefits in providing low latency [8].

The drawback of these fog nodes is their limited capacity. They cannot provide high computational and storage power [9]. Because of this reason, some data is offloaded from these fog nodes to the cloud data center for processing. ‘

Fig 2. Explains the architecture of fog and cloud. From the figure, we can observe that there are three levels. At the lowest level, we have sensors and actuators. Sensors gather the data from the surrounding, and this data is in the form of data streams known as tuples.

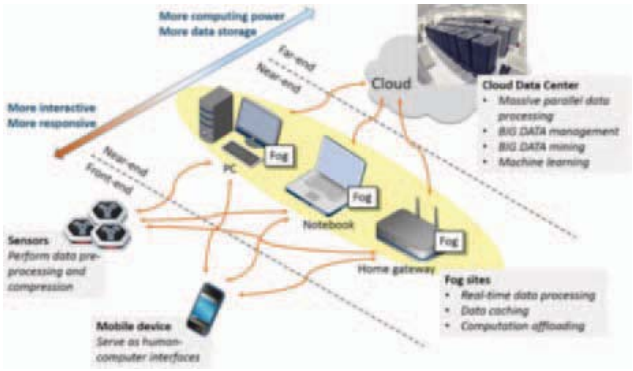


Fig. 2. Cloud/Fog Computing Architecture [10]

While actuators receive that output data from the upper layer, these sensors and actuators work as the front end. This layer is in the shortest distance with the user. This data is then transferred from this front end to near end devices, a sensor to the fog nodes, which are available in the fog layer (middle layer). These devices process the data according to the requirements and send the result back to the actuators.
If the data is not fully processed, the processed data will be offloaded to the cloud for further processing and storage purposes. Cloud may lead to a delay in the application result but will complete the job.
After completing the task, the result will be sent back to the actuators. This cloud layer is generally used for handling big applications and also for the storage purpose.
There is a various application where fog is integrated with cloud and provide support for various IoT applications very effectively like in the smart home, smart transportation, smart healthcare sytem, visual security, education sector, etc.

Aim of this is to provide a better computing experience, which can support all types of applications, even the latency-sensitive one too.
Some of the advantages of fog computing over cloud computing are low power consumption, low latency, bandwidth saving, decentralized nature, security, mobility, etc.

Location awareness is also high in fog computing in comparison to the cloud. Physically fog nodes are decentralized and distributed, while the cloud is centralized. It gives an upper edge to the fog computing and fog nodes. They can be well distributed to decrease the load on any fog node.

These advantages encourage the user to move to fog computing. As the fog layer is sharing part of the original data which it received from the terminal nodes, it is decreasing the overall bandwidth requirement, which is very useful [12].

Fog computing has some disadvantages also. These are computation and storage power of fog nodes that are much less than the cloud datacenter. But in a real environment, it does not affect much.
As the original purpose of fog computing is to provide small computation and storage facility to the end-user. So heavy computational requirement is not needed to the end-user.

In offloading, the task is to contract out to the external agents to handle. External resources work towards the completion of the intended task [13]. It may be due to the high computational requirement or storage requirement, which task source or mobile device is unable to perform.

The decision when to offload the data to the cloud is an important one. It can affect the latency significantly. If the decision when to offload the data is not taken place in time, it can degrade the whole computing experience and can affect the latency in a very negative matter.

If we upload the data to the cloud, when it was not needed, latency will be increased drastically. If we don’t upload the data to the cloud or upload late, then also the latency will be increased. We have worked in this direction and proposed a new algorithm to solve this offloading issue.

III. OFFLOADING APPROACH

A. Problem description

We are assuming that there is only one cloud datacenter, which will provide all the heavy computing facilities. It is located at a far distance from the end-user. Bandwidth between cloud and fog nodes (B_{FC}), cloud datacenter, and terminal nodes/ end devices (B_{EC}) are already known in Mbps. The energy required to transfer each unit of data from fog node to cloud datacenter (E_{FC}) and from end devices to cloud datacenter (E_{EC}) is also already known.

There is N number of fog nodes in an area with each having different processing capacity (P_N) and different pricing for each data unit (FP_N). Pricing for cloud data center varies according to the chosen VM. So, the task by end devices can be

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:27 UTC from IEEE Xplore. Restrictions apply.

configured as $T_N = \{T_1, T_{2,\ldots,} T_N\}$. It will be decided whether the fog nodes are chosen or cloud data center or both.

Bandwidth requirement to offload the data to the cloud is generally greater than the bandwidth requirement to offload the data to the fog. Also, the prices to lease the cloud datacenter resources and their capacity is much larger than the prices and capacity of fog nodes.

Total work is done in three phases. In phase 1, it is decided that whether the generated task by an end device can be processed by the end device itself or not.

If the end device is not capable of doing so, only in that condition, the decision to process the data through fog nodes will be taken. Otherwise, offloading will not be performed.

Now, if the fog node can process the data and also within the given allotted time, then the task will be processed by the fog node. Otherwise, data will be offloaded to the cloud datacenter.

There will be a central fog manager, who will be aware of all the fog node processing power and their impact zone, where they provide services.

Fog manager will select the best fog node to provide service to mobile devices based upon its computation capability and its impact zone. Tasks that fog node can not process on their own will be offloaded to the cloud datacenter.

To determine whether offloading the task to fog node or cloud from mobile devices are in our favor or not, we have proposed Mobile to Fog and Cloud Transfer (MTFCT) algorithm.

B. Execution Time

Execution time is divided into many phases, which are the following:

1) T_{OD} = A decision is made before transferring data to the fog node, whether the fog nodes can process the given task or not. If the fog nodes meet the task requirement, the task will be given to the fog nodes; otherwise, the task will be given to the cloud datacenter.
This decision is based on the task processing requirement and fog nodes capacity in the vicinity of the fog node manager. This decision takes time, and it is known as $T_{OD.}$ It is a crucial decision. It will take place with the help of various other parameters.

2) T_{MF} = Task size*Time required to transfer one unit of data to the fog (Time required to transfer whole data from mobile device to the fog node)
The selection of fog node will be managed by the fog manager. This time will only be considered when the fog manager decides that the given task can be processed by the fog node in the given time. It depends upon the available bandwidth and task size.

3) T_{MC} = Task size*Time required to transfer one unit of data to the cloud data-center (Time required to transfer whole data from mobile device to the cloud data-center). This time is generally greater than $T_{MF.}$

4) T_{EF} = Task size/fog node processing capacity (Time required to execute a task by fog node). If the task is processed by the fog node, it is calculated using the task size and processing capacity of that particular node. Lesser the processing capacity higher will be the $T_{EF.}$

5) T_{FM} = Time required to send the result back to the mobile device. The same channel is used as in T_{MF}. This time will depend upon bandwidth and result in size. The more the task size more will be its value.

6) T_{EC} = Tasksize/cloud datacenter processing capacity If the task is processed by the cloud datacenter (Time required to execute a task by cloud datacenter). It is calculated using task size and processing capacity of that cloud datacenter, which is generally very high. Its value is always less than $T_{FM.}$

7) T_{FC} = Time required to send the task from fog to cloud. Sometimes a task is partially processed by a fog node and is sent to the cloud data center for further processing. In this, the remaining task size and bandwidth between fog node and cloud datacenter play an important role. So this time is known as $T_{FC.}$

8) T_{CF} = After the task processing, if the result is sent back to the fog node, the same channel is used as in $T_{FC.}$ So this is known as $T_{CF.}$

9) T_{CM} = After the task processing, if the result is sent back to the mobile device, the same channel is used as in T_{CM}. This time will depend upon bandwidth and result in size.

10) T_M = Execution time if the task runs on mobile devices itself.

So total execution time a task requires can be anything based upon a combination of the factors mentioned above. If the cloud data center gets involved, the execution time will get high; otherwise, if only fog nodes are involved, the value of execution time will be very less.

C. Cost of Execution

We can divide our cost of execution on a remote center into two parts; the cost of execution on cloud (C_C) and the cost of execution on fog (C_F). Cost of execution on fog can be further divided into two parts which are following;

$$C_F = C_{EF} + C_{EC} \quad (1)$$

(C_{EC} depends upon how much part of the task is sent for processing by the local fog node to the cloud node)
Where C_{EF} is a cost for executing the part of the task on fog node only, and C_{EC} is the cost for the part sent to the cloud datacenter from fog node for further processing and other purposes. If the task is processed fully by the fog nodes themselves, the value of C_{EC} will become zero. So in that case,

$$C_F = C_{EF}$$

C_M = cost to run the task on mobile devices itself.
If we don't offload our task to the cloud or fog node, in that case, we calculate this value. This is generally used to check whether it is beneficial to offload the task or not. If it is not beneficial, the task will not be offloaded. But it is not the only

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:27 UTC from IEEE Xplore. Restrictions apply.

factor which decides; other factors are also there like execution time, energy requirement, etc.

Generally, the cost of processing on cloud takes high value, but still, some tasks can not be processed locally by fog nodes or the mobile devices itself, fully because of its size and required time.
So, in that case, offloading the task to the cloud datacenter becomes necessary. C_{EC} will also be affected by the scheduling policy chosen for its task scheduling.

D. Energy Requirement

Energy requirements can also be divided into two parts; E_F and E_C.

$E = E_F + E_C$. (2)

$E_F = E_{MF}+E_{EF}+E_{FC}+ E_{EC} +E_{CF}+ E_{FM}$ (3)

$E_C = E_{MC}+E_{EC}+E_{CM}$ (4)

Where,

E_F = Total energy required if task go to the fog nodes for processing. Task can also be partially executed on the fog nodes too and rest can be sent to cloud. In this case value of E_F will get vey low, because only a part of task is executing on fog node irrespective of full task.
The value of total energy consumption (E) in equation number two becomes zero if a task is decided to be processed by the end device (mobile device) itself.
E_{MF} = Task size*Energy required to transfer one unit of data (Energy required to transfer whole data to fog node)
E_{EF} = Task size*Energy required to process one unit of data by fog node (Energy required to process whole data by fog node)
E_{FC} = Energy required to transfer the remaining task for further processing by the cloud datacenter. Its value will become zero if the task is fully processed by the fog nodes themselves.
E_{CF} = Energy required to transfer the result to fog node by cloud. Its value will become zero if the task is fully processed by the fog nodes themselves.
E_{FM} = Energy required to transfer the result to the mobile device by fog node
E_C = Total energy required if the task goes to the cloud data center for processing. Task can also be partially executed on the cloud datacenter too. Its value will become zero if the task is fully processed by the fog nodes themselves.
E_{MC} = Task size*Energy required to transfer one unit of data to the cloud datacenter (Energy required to transfer whole data to the cloud datacenter). This is generally higher than the E_{MF} as the cloud data center is generally very far located as compared to the fog nodes.
E_{EC} = Task size*Energy required to process one unit of data by cloud datacenter (Energy required to process whole data by cloud datacenter)
E_{CM} = Energy required to transfer the result to the mobile device by cloud datacenter.
E_M = Energy required to run the task on the mobile device itself

E. Proposed algorithm

Algorithm 1 MTFCT algorithm
Input: parameters like T_{OD}, T_{MF}, T_{EF}, T_{FC}, T_{EC}, T_{CF}, T_{FM}, T_{CM}, T_{MC}, T_M.
Output: Offloading decision, wheather to offload to fog, cloud or not

1: Calculate the value of each variable like T_{OD}, T_{MF}, etc.
2: Check how much part of the task can be executed on local fog node too.
3: Calculate $T_F = T_{OD}+T_{MF}+T_{EF}+T_{FC}+T_{EC}+T_{CF}+T_{FM}$ (T_{EC} depends upon how much part of task is sent to cloud for processing) (5)
4: Calculate $T_C=T_{OD}+T_{MC}+T_{EC}+T_{CM}$ (6)
5: Compare T_M with T_F and T_C.
  5.1: If $T_M > T_F$ and $T_M < T_C$, Task will be uploaded to the fog node
  5.2: If $T_M > T_F$ and $T_M > T_C$, check if $T_F > T_C$, If yes Task will be uploaded to the cloud data center otherwise to the fog node
  5.3: If $T_M < T_F$ and $T_M > T_C$, The task will be uploaded to the cloud datacenter
  5.4: If $T_M <= T_F$ and $T_M <= T_C$, The task will not be offloaded to either fog node or cloud datacenter
6: Similarly, compare C_M with C_F and C_C
7: Also, calculate E_F and E_C
8: Compare it with E_M
9: Based upon the above comparisons and the weighted value of each factor (execution time, cost & energy), we decide to offload the task from mobile device to fog node or cloud node is beneficial or not take decision based upon that. (Based upon how much value or importance is given to each factor by user)

IV. Conclusion

Fog computing is the lifeline of IoT in today's world scenario. IoT can be implemented without fog computing, too, but with the help of fog computing, the efficiency of IoT increases considerably. So to achieve very high efficiency, we integrate fog computing with cloud computing.

The energy requirement of fog is very less as compared to the energy required to process the same task in a cloud datacenter. This can help us in reducing the total co2 emission to save the environment too.

In this paper, we proposed a method for task offloading to fog nodes and cloud datacenter using various parameters. In this, the involvement of users is also very high.

This is a theoretically proposed algorithm, so in the future, we would like to implement this proposed algorithm in a real fog computing environment or in a simulated environment using FogSim simulation toolkit and also provide node to node migration of task in fog computing environment and along with a better way to handle the task on cloud datacenter whether if the task is offloaded from a mobile device or any fog node

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:27 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] S. Zahra, M. Alam, Q. Javaid, A. Wahid, N. Javaid, S.U.R. Malik and M.K. Khan, "Fog computing over IoT: A secure deployment and formal verification," IEEE Access 5 (2017): 27132-27144.

[2] M.A. Rodrigue and R. Buyya, "A taxonomy and survey on scheduling algorithms for scientific workflows in IaaS cloud computing environments," Concurrency and Computation: Practice and Experience, 29(8), 2017.

[3] H.A. Khattak, H. Arshad, S.U. Islam, G. Ahmed, S. Jabbar, A.M. Sharif and S. Khalid, "Utilization and load balancing in fog servers for health applications." EURASIP Journal on Wireless Communications and Networking 2019.1 (2019): 91.

[4] S. Khan, S. Parkinson and Y. Qin, "Fog computing security: a review of current applications and security solutions," Journal of Cloud Computing 6.1 (2017): 19.

[5] D. Poola, M.A. Salehi, K. Ramamohanarao and R. Buyya, "Fog computing: A taxonomy, survey and future directions," Internet of Everything (pp. 103-130). Springer, 2016.

[6] C. Puliafito, E. Mingozzi, F. Longo, A. Puliafito and O. Rana, " Fog computing for the Internet of Things: A Survey," ACM Transactions on Internet Technology, Vol. 19, No. 2, Article 18, 2019 .

[7] A. Mohammad, and E. N. Huh. "Fog computing micro datacenter based dynamic resource estimation and pricing model for IoT." 2015 IEEE 29th International Conference on Advanced Information Networking and Applications. IEEE, 2015.

[8] S. Delfin, S.P. Sivasanker, N. Raj and A. Anand, "Fog computing: A new era of cloud computing." 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019.

[9] F. Haouari, R. Faraj and J.M. Alja'am, "Fog computing potentials, applications, andcChallenges," 2018 International Conference on Computer and Applications (ICCA). IEEE, 2018.

[10] J. K. Zao, T. T. Gan, C. K. You, C. E. Chung, Y. T. Wang, S.J.R. Méndez, et al., "Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology." Frontiers in human neuroscience, 8, 370, 2014.

[11] R. Jindal, N. Kumar, and H. Nirwan, "A survey on cloud to fog evolution" unpublished.

[12] Y. Sun, and N. Zhang, "A resource-sharing model based on a repeated game in fog computing." Saudi journal of biological sciences, 24(3), 687-694, 2017.

[13] M. Aazam, S. Zeadally, and K. A. Harras. "Offloading in fog computing for IoT: Review, enabling technologies, and research opportunities." Future Generation Computer Systems 87: 278-289,2018.

[14] A. Khakimov, A. Muthanna, and M. S. A. Muthanna. "Study of fog computing structure." IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). IEEE, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:27 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020)
IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2

NLP based Machine Learning Approaches for Text Summarization

Rahul
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
Rahul@dtu.ac.in

Surabhi Adhikari
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
Surabhi_bt2k18@dtu.ac.in

Monika
Department of Computer Science and Engineering
Delhi Technological University
New Delhi, India
Monika.siwaliya@gmail.com

2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) 978-1-7281-4889-2/20/$31.00 ©2020 IEEE 10.1109/ICCMC48092.2020.ICCMC-00099

***Abstract*— Due to the plethora of data available today, text summarization has become very essential to gain just the right amount of information from huge texts. We see long articles in news websites, blogs, customers' review websites, and so on. This review paper presents various approaches to generate summary of huge texts. Various papers have been studied for different methods that have been used so far for text summarization. Mostly, the methods described in this paper produce Abstractive (ABS) or Extractive (EXT) summaries of text documents. Query-based summarization techniques are also discussed. The paper mostly discusses about the structured based and semantic based approaches for summarization of the text documents. Various datasets were used to test the summaries produced by these models, such as the CNN corpus, DUC2000, single and multiple text documents etc. We have studied these methods and also the tendencies, achievements, past work and future scope of them in text summarization as well as other fields.**

***Keywords*—** *Natural Language Processing(NLP), Machine Learning(ML), Neural Network(NN), Abstractive(ABS) and Extractive(EXT) method*

I. INTRODUCTION

Today's world is centralized on computers and data. Data are our intangible thoughts and imagination. We are producers and consumers of data at the same time. Every little thing in our mundane lives are either a source or receiver of data. For, example when we drive there's data involved, the speed of the car, mileage, distance traveled, etc. Since 20th century, data has been a significant part of our lives, but these days we infer more from data. We store and access them through electronic and wireless systems.

Since the advent of the internet, there's an enormous amount of data available today. The Internet is a storehouse of data. Information on news, movies, education, medicine, health, nations, weather, geology, etc. is available on the internet. This could be statistical, numerical, mathematical or text data. Text data is more difficult to interpret due to larger amount of characters. Due to this gigantic amount of information, there must be a system in order to get only the essential parts of the information we access. Text summarization is a way of doing this.

Text summarization has been a topic of research and study since decades. Various models have been proposed and tested on different datasets to generate concise summaries. They are compared with different comparison scores. Text summarization can be EXT or ABS, single document or multi-document, and query-based or generic.

EXT text summarization is a way of generating summaries by using the same sentences as in the document. ABS is more general and focuses on key concepts of the document. Similarly, single document summarization techniques give summaries of the text of a single document, and multi-document generates summaries of multiple documents. Moreover, these days, there's a need for summarizing text based on queries. Query-based summarization models give summaries of the text based on a specific area as described by the query given by the user, whereas generic summaries are mostly ABS that focus on the general area of the text input.

Text summarization has been extensively used in various fields like science, medicine, law, engineering, etc. Researchers have focused on generating summaries of doctor's prescriptions, and that has been proved very useful to patients. Similarly, long news articles have been summarized and this way readers can gain a lot of information on several topics within a short span of time[1]. In this paper, we have discussed the various methods used in text summarization for past five years. The most common methods were found to be Machine Learning(ML), NNs, reinforcement learning, sequence to sequence modeling and fuzzy logic. Similarly, various optimization methods have been used to optimize the proposed objective function for the purpose of text summarization. We can see that various methods were tested on the same dataset, and their accuracy scores were found to be different. Moreover, we can also see that some researchers have combined the different methods and found out the summaries are more accurate than when a single method is used. When NLP processing has been used as a technique to summarize text documents, we see that python libraries such as scikit learn, nltk, spacy, and fastai has been used.

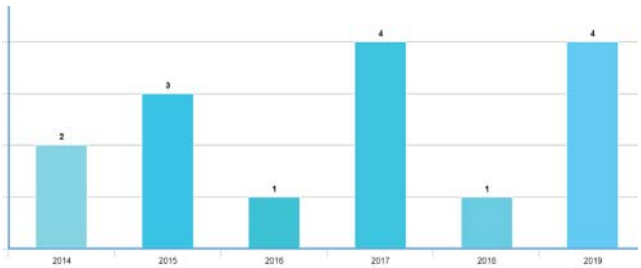


Fig. 1. Distribution of papers studied over the years

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:36:33 UTC from IEEE Xplore. Restrictions apply.

II. Related Work

Massimo Mauro, et al. have used a sentence extraction method to generate EXT summaries. In this method, the sentences are checked for their relevance and scored accordingly. Similar sentences were then clustered together to discover the most informative sentences and they were selected on the basis of sentence scores[2].

Sarda A.T. et al. have proposed NNs and Rhetorical Structure Theory (RST) to produce EXT summaries of articles. Features are compared and combined to produce a relevant sentence for the summary. The first step is to train the NN so that it learns to select the types of sentences that would be present in the summary[3]. After that, sentences in the document are selected and checked if they fit into the summary. After finding these sentences, they are fed to the rhetorical structure with the help of which linguistic relations are distinguished to form a better summary.

Gabriel Silva et al. have performed experiments on CNN-Corpus to generate EXT summaries of the documents. . . The sentences of the corpus after removing figures, videos, and tables were assigned with feature vectors for scoring. The dimensionality of feature vectors was reduced using the selection algorithms of WEKA viz. CFS Subset Evaluator, Information Gain Evaluator, and SVM Attribute. Naive Bayes was proven to be the best classifier among 5 classifiers tested using WEKA's platform[4].

Taeho Jo has presented an approach to summarize text by considering similarity between attributes and using KNN algorithm. A paragraph is given as input and it is divided into sentences. Words are represented as vectors. Each sentence is then classified as summary or remaining, and based on the similarity score between them and a human generated summary, it's included in the summary[5]. In this paper, feature is considered as an important factor in order to summarize a text paragraph. This approach can be used in multiple areas like medicine, law, engineering[5]. Similarity between only some features are taken into account, and similarity score is calculated using the KNN method, based on which a paragraph is summarized.

Mahsa Afsharizadeh, et al. presented a query-based approach for EXT text summarization. It selects the relevant sentences from the document and includes it in the summary. Eleven query-dependent appropriate features were chosen and used in the paper to find the important sentences[6]. For finding the useful sentences in the document, each sentence is assigned a score by checking on the linear function of its feature values.

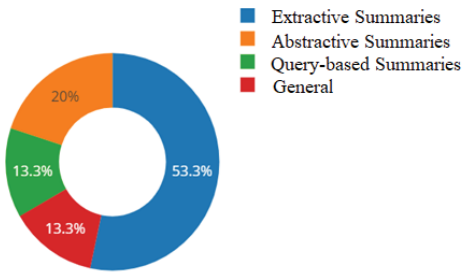


Fig 2. Types of Summaries Studied

For training and testing purpose, The DUC 2007 corpus was used. This method was able to get better average precision, average recall, and average F-score than earlier methods compared in the paper[7].

TABLE I. Summary of the Papers Analyzed

Name of Author	Year of Publication	Methods Used	Dataset used	Remarks
Kamalanathan Kandasamy et. Al[9]	2014	URL analysis, NLP, and Supervised, ML Techniques.	Tweets from users.	Combining three models gives more accuracy than single method used.
Mohamed Abdel Fattah [10]	2014	k—means clustering, differential evolutionary algorithm	data sets from DUC2001 and DUC2002	Results were compared with ROGUE 1 and ROGUE 2 scores. It showed better results than other text summarizers. Graph based algorithms for clustering could improve the result.
Mr. Sarda A.T. et al [3]	2015	NN, ML, RST	Input text documents	Due to inclusion of numerical data features, the output of this summarization method is better than other online text summarizers.
Gabriel Silva et al[4]	2015	STOM, CSF, Information Gain Evaluator and SVM attribute	news texts from CNN corpus	The results confirmed ML techniques improves the text summarization results.
K. Vimal Kumar, et al[15]	2015	Sentence ranking, Sentential semantic analysis and Sentence extraction	Hindi-text Single documents	This method can be applied for multi documents.
Rasim Alguliyev et al[12]	2016	Human Learning Optimization Algorithm	Takes a document as input.	This model tries to find balance between the topic coverage and sentence repetition in a summary, for generating concise summary.
Taeho Jo[5]	2017	k-nearest neighbor algorithm	Input paragraph	Similarity score is taken into account. Can be used in medicine, science and law.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:36:33 UTC from IEEE Xplore. Restrictions apply.

Massimo Mauro et al.[2]	2017	sentence extraction, sentence scoring and sentence ordering using Python libraries such as scikit learn, nltk.	DUC2001 dataset	It was mostly useful for multi-text summaries when compared with ROUGE score
Aditya Jain, Divij Bhatia et. Al[14]	2017	Word vector embedding, NN	100 news articles from CNN news with their ABS summaries.	the proposed model outperformed other online text summarizers such as SPLITBRAIN, AutoSummarizer when the ROUGE scores were compared.
Nicholas Giamblanco et al[11]	2017	Newtonian method	SemEval from the University of Waikato	The method outperformed RAKE and Tf-IDF scores.
Mahsa Afshariadeh, et al[7]	2018	Query based EXT summarization using TF-IDF, fuzzy logic and LSA.	DUC 2007 corpus	A fluent 250-word summary is generated and compared to ROGUE score.
J.N.Madhuri et al[8]	2019	Sentence ranking method using Python 3.6 and NLTK[8]	Takes text document as input	Those sentences whose rank is greater than 8 are included in the summary. The summarized text is then converted to audio form.
Milad Moradi et al[13]	2019	BERT model	Test on preliminary experiments of the development corpus	The size of summary would be good if it is 30% of the original text. This method worked best on biomedical texts.
Begum Mutlu et al[6]	2019	ML, fuzzy network	Input text documents and their features.	The resulting rule sets provide more coverage and generalization
Jesus M. Sanchez-Gomez et al[1]	2019	Multi Objective Artificial Bee Colony.	DUC 2002 datasets.	Efficiency problems were seen in this algorithm and they were solved using the asynchronous parallel MOABC.

III. Overview of Machine Learning Methods for Text Summarization

We have seen various researchers presenting ML methods for text summarization. Various supervised ML models such as Naïve Bayes, Random Forest, SVD have been used to generate EXT text summaries. Kamalanathan Kandasamy, et al. have used ML technique to classify spams on twitter. The data sets were also tested using SVM, but Naïve Bayes was found to be more accurate. For this experiment, 100 users were tested. Out of them, 60 were legitimate and 40 were spam. 98 users were categorized correctly[9].

We have also seen many researchers use deep learning techniques for EXT as well as ABS text summarization. Deep learning is an area of ML. Various NN techniques have been used. Similarly, reinforcement learning, Convolutional NN(CNN), RNN have also been applied to generate text summaries[10]. There's also a study of sequence-to-sequence models for text summarization these days. These methods are extension of ML. We describe some of the papers that use the above mentioned methods to generate summaries of text.

Nicholas Giamblanco et al presented a Newtonian method to generate key phrases. The author has used four steps for keyword and key phrase extraction. First the stop words are removed which is known as noise filtering, then each words is assigned a mass, and the relations between words is calculated which is also known as word attraction, and finally a key phrase is generated[11]. To test the accuracy of this algorithm, it was tested against RAKE and TF-IDF scores. The data used was from SemEval from the University of Waikato[12]. The author has also shown how Key-LUG has outperformed RAKE and TF-IDF. The metrics used for this purpose was recall, precision and F1 score[13].

Aditya Jain et. al paper proposed a binary approach for EXT text summarization. The document is broken into sentences and checked if it's relevant or irrelevant to the main theme of the document. A NN is then used and tested if the sentence is to be included or not. 100 news articles from CNN news along with their corresponding ABS summaries have been used as data sets for this model. Sentences in the news article and summary are compared and those with higher similarity score between them are chosen for EXT summary[14]. The author has used vector embedding to represent each word in a sentence as 100 dimensional vectors. For testing the performance of the proposed model, first 284 documents of the DUC 2002 dataset was used. It was seen that the proposed model was more accurate other online text summarizers[15].

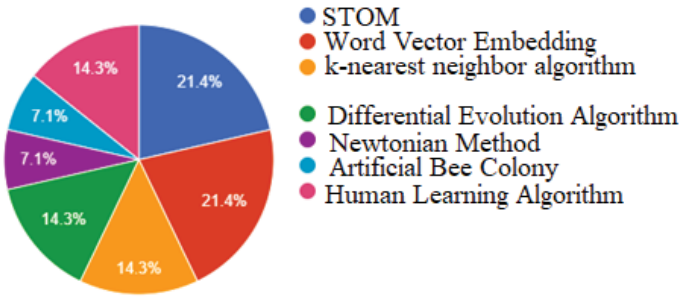


Fig 3. Different algorithms used

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:36:33 UTC from IEEE Xplore. Restrictions apply.

IV. Conclusion and Future Work

We have seen that due to abundant availability of data, text summarization has a very vital role in saving user's time, as well as resources. Text summarization is indeed an important tool for today. We have seen the use of various algorithms and methods for this purpose. These methods, in individual and together give different types of summaries. Their accuracy score can be compared to find the better and more concise summaries. For this purpose, ROGUE score has been used more frequently. Similarly, in some cases TF_IDF scores have been used too.

The summaries generated using these methods are not always up to the mark. Sometimes, it's also irrelevant to the original document. Therefore, this topic is ongoing and people have done various works on this. There isn't any specific model that generates best summaries. So, for future, the models discussed can be modified for more accurate summaries. For e.g., we could use GAN's and transfer learning. For future, this can give a way to develop and enhance further ideas for text summarization.

Acknowledgment

We are extremely grateful to the Department of Computer Science and Engineering, Delhi Technological University for providing us the platform and all the necessary materials required to carry out this study.

References

[1] C. Ordonez, Y. Zhang, and S. L. Johnsson, "Scalable machine learning computing a data summarization matrix with a parallel array DBMS," *Distrib. Parallel Databases*, vol. 37, no. 3, pp. 329–350, 2019, doi: 10.1007/s10619-018-7229-1.

[2] M. Mauro, L. Canini, S. Benini, N. Adami, A. Signoroni, and R. Leonardi, "A freeWeb API for single and multi-document summarization," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1301, 2017, doi: 10.1145/3095713.3095738.

[3] A. T. Sarda and M. Kulkarni, "Text Summarization using Neural Networks and Rhetorical Structure Theory," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 49–52, 2015, doi: 10.17148/IJARCCE.2015.4612.

[4] G. Silva, R. Ferreira, S. J. Simske, L. Rafael Lins, M. Riss, and H. O. Cabral, "Automatic text document summarization based on machine learning," *DocEng 2015 - Proc. 2015 ACM Symp. Doc. Eng.*, pp. 191–194, 2015, doi: 10.1145/2682571.2797099.

[5] T. Jo, "K nearest neighbor for text summarization using feature similarity," *Proc. - 2017 Int. Conf. Commun. Control. Comput. Electron. Eng. ICCCCEE 2017*, pp. 1–5, 2017, doi: 10.1109/ICCCCEE.2017.7866705.

[6] B. Mutlu, E. A. Sezer, and M. A. Akcayol, "Multi-document extractive text summarization: A comparative assessment on features," *Knowledge-Based Syst.*, vol. 183, p. 104848, 2019, doi: 10.1016/j.knosys.2019.07.019.

[7] M. Afsharizadeh, H. Ebrahimpour-Komleh, and A. Bagheri, "Query-oriented text summarization using sentence extraction technique," *2018 4th Int. Conf. Web Res. ICWR 2018*, pp. 128–132, 2018, doi: 10.1109/ICWR.2018.8387248.

[8] J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," *2019 Int. Conf. Data Sci. Commun. IconDSC 2019*, pp. 1–3, 2019, doi: 10.1109/IconDSC.2019.8817040.

[9] K. Kandasamy and P. Koroth, "An integrated approach to spam classification on Twitter using URL analysis, natural language processing and machine learning techniques," *2014 IEEE Students' Conf. Electr. Electron. Comput. Sci. SCEECS 2014*, pp. 1–5, 2014, doi: 10.1109/SCEECS.2014.6804508.

[10] M. A. Fattah, "A hybrid machine learning model for multi-document summarization," *Appl. Intell.*, vol. 40, no. 4, pp. 592–600, 2014, doi: 10.1007/s10489-013-0490-0.

[11] N. Giamblanco and P. Siddavaatam, "Keyword and Keyphrase Extraction using Newton's Law of Universal Gravitation," *Can. Conf. Electr. Comput. Eng.*, pp. 1–4, 2017, doi: 10.1109/CCECE.2017.7946724.

[12] R. Alguliyev, R. Aliguliyev, and N. Isazade, "A sentence selection model and HLO algorithm for extractive text summarization," *Appl. Inf. Commun. Technol. AICT 2016 - Conf. Proc.*, pp. 1–4, 2017, doi: 10.1109/ICAICT.2016.7991686.

[13] M. Moradi, G. Dorffner, and M. Samwald, "Deep contextualized embeddings for quantifying the informative content in biomedical text summarization," *Comput. Methods Programs Biomed.*, vol. 184, p. 105117, 2020, doi: 10.1016/j.cmpb.2019.105117.

[14] A. Jain, D. Bhatia, and M. K. Thakur, "Extractive Text Summarization Using Word Vector Embedding," *Proc. - 2017 Int. Conf. Mach. Learn. Data Sci. MLDS 2017*, vol. 2018-Janua, pp. 51–55, 2018, doi: 10.1109/MLDS.2017.12.

[15] J. K. Mandal, S. C. Satapathy, M. K. Sanyal, P. P. Sarkar, and A. Mukhopadhyay, "Information systems design and intelligent applications: Proceedings of second international conference India 2015, volume 1," *Adv. Intell. Syst. Comput.*, vol. 339, pp. 301–310, 2015, doi: 10.1007/978-81-322-2250-7.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:36:33 UTC from IEEE Xplore. Restrictions apply.

Noisy three-player dilemma game: Robustness of the quantum advantage

Pranav Kairon[§], Kishore Thapliyal[‡,*], R. Srikanth[♮], and Anirban Pathak[♠,†]
[§]Delhi Technological University, Bawana Road, Delhi, 110042, India
[‡]RCPTM, Joint Laboratory of Optics of Palacky University and Institute of Physics of Academy of Science of the Czech Republic, Faculty of Science, Palacky University, 17. listopadu 12, 771 46 Olomouc, Czech Republic
[♮]Poornaprajna Institute of Scientific Research, Bengaluru, Karnataka 560080, India
[♠]Jaypee Institute of Information Technology, A-10, Sector-62, Noida, UP-201309, India

arXiv:2004.04533v1 [quant-ph] 9 Apr 2020

Abstract

Games involving quantum strategies often yield higher payoff. Here, we study a practical realization of the three-player dilemma game using the superconductivity-based quantum processors provided by IBM Q Experience. We analyze the persistence of the quantum advantage under corruption of the input states and how this depends on parameters of the payoff table. Specifically, experimental fidelity and error are observed not to be properly anti-correlated, i.e., there are instances where a class of experiments with higher fidelity yields a greater error in the payoff. Further, we find that the classical strategy will always outperform the quantum strategy if corruption is higher than half.

1 Introduction

Game theory provides a way to learn about decisive communication between rational and self-seeking agents. Therefore, it plays an important role in the fields of computer science, economics, biology, psychology, etc. (see [1, 2] for review). Computationally, game theory can be used to model algorithms [3, 4] as well as check the robustness of networks and corresponding attack strategies [5]. In cryptography, the communication task can be visualized as a game between the parties trying to communicate securely and an eavesdropper ([6] and references therein). With the advent of quantum computing, it is observed that resources used in quantum computing, such as quantum coherence and entanglement, provide alternative solutions to classical games.

We may mention, for example, the emergence of cooperation in the prisoner's dilemma game [7] and the resolution of the coordination in battle of sexes game [8] using entanglement. Specifically, as all the players wish to maximize their gain or payoff in games, for which the umpire has laid down the rule(s), players using quantum mechanical tactics are found to attain a higher payoff compared to the classical one [9]. Further, the dilemma disappears in prisoner's dilemma with the use of quantum resources under unitary operations [10, 11]. Along the same line, optimal cloning of quantum states is also studied as game [12]. Quantum games based on monogamy of entanglement are shown to be useful in device independent quantum cryptography [13]. Our understanding of several other foundational aspects of quantum mechanics is improved by considering games, such as nonlocality [14], the uncertainty bound on nonlocality [15], contextuality [16], PR-boxes [17], as well as applications in quantum reinforcement learning [18] and quantum machine learning [19].

Over the course of time, multiplayer quantum games were also introduced that exploit quantum correlation to prevent betrayal by individual players [20]. It has been suggested that these quantum games may shed light on the interactions in many-particle systems [21]. One such multiplayer game is the three-party counterpart of the prisoner's dilemma. In the classical version, all three players prefer to choose analogous to the corresponding two-party case. The dilemma exists because the Nash equilibrium does not coincide with the Pareto optimal [20]. Specifically, a Nash equilibrium is the situation in which no participant can gain by a unilateral change of strategy, while Pareto optimal corresponds to the situation that any change in strategy would make at least one individual worse off [20]. Still in quantum case, use of tripartite entanglement shows certain advantage. Moreover, computing the Nash equilibrium in the three- and four- player games is shown to be a hard problem [22, 23]. An experimental verification of three-player dilemma game using NMR was reported in Ref. [24]. In the recent past, other games have been realized on photonic quantum computer [25–28] and ion trap platform [29].

In general, the dilemma games are relevant in several studies of biology, economics, psychology, international relations, sports to name a few. For instance, King Solomon's dilemma [30] based on the Old Testament can model prize

*Email: kishore.thapliyal@upol.cz
†Email: anirban.pathak@jiit.ac.in

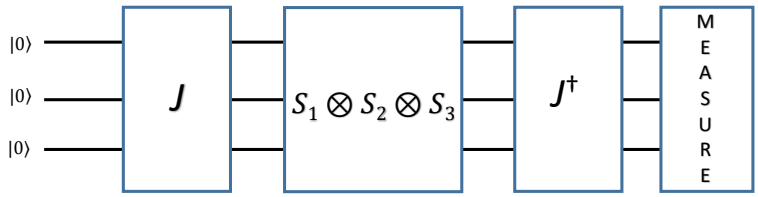


Figure 1: (Color online) Circuit diagram that can be used for the realization of the three-player quantum dilemma game.

allocation, research grant distribution, etc. Another multiparty version of prisoner's dilemma is diner's dilemma in which each player has to choose whether to order an expensive or an inexpensive dish if they have to equally share the bill [31]. This iterated diner's dilemma is useful in the social dynamics of networks and situational awareness. Such iterated multiparty prisoner's dilemma in the context of social dynamics is discussed in the past, too [32]. Along the same line, dilemma of the players in other games is used to introduce the conditional probability [33].

Decoherence is the Achilles' heel of quantum computing and information processing in particular, and technology in general. Similar results are shown for the quantum games [34]. Independently, the effect of errors in the initial state preparation (as corruption by a demon) on the outcome of three-player dilemma game is studied assuming that the players are unaware of corruption and that there is no decoherence [21]. Interestingly, beyond a pivotal value of corruption it can be observed that players fare off better with the classical strategies, but since players have no knowledge of the level of corruption they have to stick to their original strategies. Furthermore, a quantum game reduces to classical game if one of the parties allows his qubit to decohere under Markovian noise channels [35], while Nash equilibria are unchanged by decoherence for prisoner's dilemma [36].

Here, we wish to implement the three-player dilemma game [20] on IBM quantum computer [37] and study how the change in the utility function affects the point of quantum advantage. Interestingly, this is the first realization of a game with corrupt source on a superconducting qubits based quantum computer. Despite high error rate and the limited qubit connectivity, it has been shown to run a wide array of algorithms ([38, 39] and references therein). Thus, we realize the game on IBM Q Experience and compare the experimental payoffs with previous experiments on NMR [24]. On generalizing the payoff table in the noisy game, the point where quantum advantage disappears also changes which leads to some interesting observations. In specific, they show how robust the quantum strategy is. An application of these results is that given a known corruption level, the payoff table (the relative stakes) may, in a range, be chosen to give an advantage to the quantum strategy. Finally, we show that classical strategies dominate when corruption is higher than 50% in the proposed game.

The rest of the paper is organized as follows. We introduce three-player quantum dilemma in Section 2. The noisy counterpart of the game and its experimental implementation is discussed in Section 3. We further discuss all the results in detail in the penultimate section before concluding the paper in Section 5.

2 Three-person quantum dilemma

A multiparty dilemma game was introduced as a multiparty counterpart of the prisoner's dilemma game, where each person has two choices: either to cooperate (0) or defect (1). The three-player dilemma resembles el-Farol bar problem that players have to decide independently whether to go or not to a bar with seating capacity for only two ([24] and references therein).

In the three-player quantum dilemma game, each player is provided one qubit by the umpire, who performs an entangling operation on state $|000\rangle$ before that which increases the nonclassical correlation among the players. The entangling gate J can be defined as in Ref. [20]:

$$J = \cos\frac{\gamma}{2} I^{\otimes 3} + i \sin\frac{\gamma}{2} X^{\otimes 3}, \tag{1}$$

where I and X are identity and Pauli NOT gates, respectively. Without loss of generality, we choose the case when the correlation is maximum, i.e., $\gamma = \pi/2$. Further, it can be checked that for minimum correlation, i.e., $\gamma = 0$, the game reduces to its classical counterpart [11].

In quantum game, each person is allowed to choose an operation from a strategy set S, consisting of 3 elements $S = (S_1, S_2, S_3)$, where $S_1 = X$ means the player wants to attend the party; $S_2 = H$ corresponds to the player's choice to go with half a probability; and $S_3 = I$ represents the player wants to stay at home. Note that the choice of S_2 does not have a counterpart in classical games. This is a restricted strategy set (as there can be an infinitely many possible quantum strategies each corresponding to a different unitary operation), but it encompasses all the nonclassical characteristics we want to demonstrate through this game. Subsequently, a disentangling operation $J^\dagger = \frac{I^{\otimes 3} - iX^{\otimes 3}}{\sqrt{2}}$ is performed before measuring in the computational basis. The circuit diagram of the game is shown in Fig. 1.

Thus, when none of the players decide to go, i.e., the measurement outcome is $|000\rangle$ (represented by corresponding bit values 000 in Table 1), nobody is happy since they could not attend the party but are not sad since none of the friends

Bit values corresponding to possible measurement outcome	Payoffs ($)
000	$0, 0, 0$
001	$-n, -n, p$
010	$-n, p, -n$
011	p, n, n
100	$p, -n, -n$
101	n, p, n
110	n, n, p
111	q, q, q

Table 1: Generalized payoff table for three-player dilemma game depending upon the possible measurement outcomes. In previous adaptations of the game [20], $n = 9$, $p = 1$, and $q = 2$ were used.

betrayed, and thus everybody gets 0 payoff. However, if one person decides to go then the other two will be unhappy (with payoff $-n$), and the one attending the party does not enjoy being alone (with payoff p). When two of the friends decide to go they both fare off with n payoff each since they get to go to the party with company, while the friend left behind is not too dejected since his presence would have overcrowded the party so he gets p. If all of them decide to go, they get payoff of q each since their presence has overcrowded the party. Accordingly, we impose $n > q > p > 0$. This problem is also relevant in the recent pandemic coronavirus situation that may allow a restaurant to open but to avoid infection it restricts people who can sit at a table to two, say because the table is 1 m wide and only two persons can sit in the same table opposite to each other without violating the social distancing norms.

Yet another example for the three-player dilemma closer to most of us would be the dilemma of three academic collaborators in applying for a research grant. If two of them apply, they are likely to receive the grant, whereas they probably would receive insufficient or no funding if all three apply for it. Also, they would not be happy if none of them apply or their collaborator gets it but not them. The dilemma shown previously was by considering $n = 9$, $p = 1$, and $q = 2$ [20]. Here, we study a general description of such payoff tables and show how the payoff depends on these parameters in the noisy game (subjected to constraint $0 < p < q < n$). One motivation for this is to understand whether and how the game's stakes can be fixed based on knowledge of the preparation noise in the system. In [20], it is shown that in a special case considering a different set of values of the payoffs, quantum players do not have any advantage over the classical strategy which is no longer a Nash equilibrium. Therefore, we have restricted ourselves to the aforementioned constraint which ensures that classical Nash equilibrium exists. To the best of our knowledge, this is the first attempt to generalize the payoff table for the three-player dilemma game in analogy of prisoner's dilemma [40].

3 IBM implementation and noisy state preparation

For the given strategy space S, there are 3 choices per player which gives us $3^3 = 27$ arrangements which can be clustered into 10 different classes [21]. Experimental design for the implementation of all these classes is shown in Fig. 2. Classes I, IV, and V have $\binom{3}{3} = 1$ configuration, while Classes II, III, VI, VII, IX and X each have $\binom{3}{1} = 3$ possible configurations, and there are $3! = 6$ configurations in Class VIII. When each player decides to play a unique tactic, independent of the other players, we obtain class VIII as mixed strategy Nash Equilibrium. Class VII provides best response for each player but is not considered as it is biased. Suppose the umpire has provided tainted qubits from the black box, i.e., instead of $|000\rangle$ he introduces error (mixedness) of the form $(1-x)\,|000\rangle\langle 000| + x\,|111\rangle\langle 111|$. The expected payoff would change quite exorbitantly as we increase the amount of preparation noise or corruption (as shown in Fig. 6), where 'x' is the noise parameter, $x \in [0, 1]$.

In IBM implementation [41] of the gate [42], we performed the entangling gate J using a $R_x(-\frac{\pi}{2})$ and 4 CNOT gates as

$$J = (I_0 \otimes CNOT_{1\to 2})\ (CNOT_{1\to 0} \otimes I_2)\ \left(I_0 \otimes R_x\left(-\frac{\pi}{2}\right)_1 \otimes I_2\right)\ (I_0 \otimes CNOT_{1\to 2})\ (CNOT_{1\to 0} \otimes I_2)$$

on qubits 0, 1, and 2, respectively (also shown in the Fig. 3). The single qubit operation can be defined as $R_x(-\frac{\pi}{2})\,|0\rangle \to (|0\rangle + i\,|1\rangle)/\sqrt{2}$. In fact, many simulations of this game have modeled the entangling gate J incorrectly in the past [42] as those matrix decomposition were not the same as J. As already discussed this is followed by the players applying their operations on their respective qubits from the strategy space S. Finally, we apply a disentagling gate $J^\dagger$ which can be modeled by the same gates as J since $(ABCDE)^\dagger = E^\dagger D^\dagger C^\dagger B^\dagger A^\dagger$.

To introduce the corruption in the input qubits (shown in Fig. 4), the umpire uses an ancilla in state $|\psi\rangle = U(\theta, \phi, \lambda)\,|0\rangle = \cos\frac{\theta}{2}\,|0\rangle + \sin\frac{\theta}{2}\,|1\rangle$ prepared using single qubit unitary operation $U(\theta, \phi = 0, \lambda = 0)$ defined as

$$U(\theta, \phi, \lambda) = \begin{pmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i\lambda+i\phi}\cos\frac{\theta}{2} \end{pmatrix}. \tag{2}$$

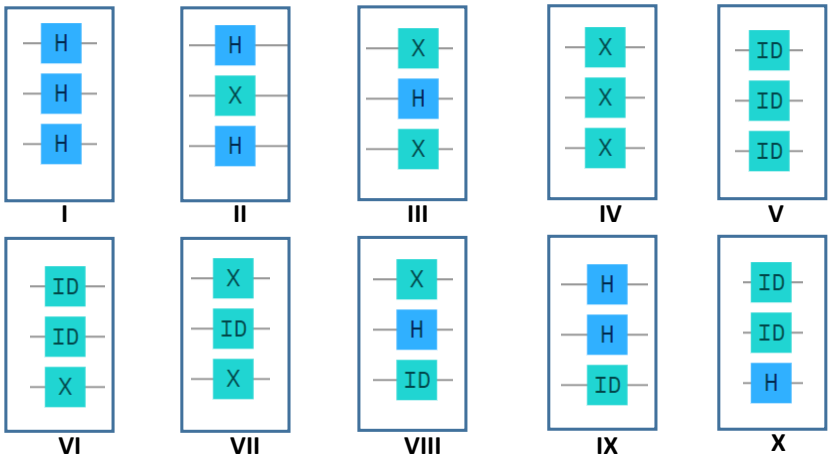


Figure 2: (Color online) Ten classes with different operations used by each player in $S_1 \otimes S_2 \otimes S_3$. Here, H, X, and ID are Hadamard, NOT, and identity gates, respectively.

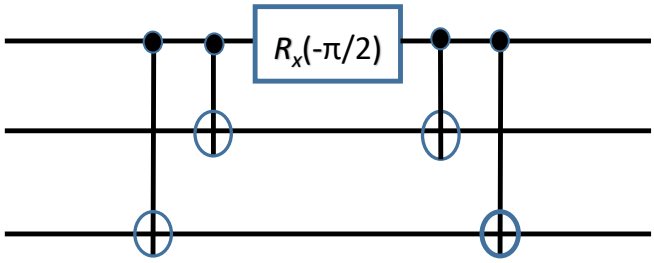


Figure 3: (Color online) Decomposition of three qubit operation J.

Subsequently, he uses this ancilla as control and applies CNOT gates to the rest of the qubits which he sends to three players. The amount of corruption, x is related by $x = \sin^2 \frac{\theta}{2}$. Therefore, in what follows, we trace out the fourth qubit to obtain the payoffs of noisy game.

4 Results and discussion

We have performed the experiments for all classes and computed payoff from the measurement outcomes in computational basis. We have also obtained the output density matrices to obtain the fidelity between theoretically desired and experimentally reconstructed states.

4.1 Fidelity and quantum state tomography

In quantum computation, fidelity is used to describe closeness between two states as it is one of the distance based measures. Ideally, fidelity between the experimental (ρ^E) and theoretical (σ) density matrices, defined as $F(\rho^E, \sigma) = Tr\sqrt{(\sqrt{\sigma}\rho^E\sqrt{\sigma})}$ [39], is desired to be 1, but due to unavoidable errors it is usually less than unity in most cases.

In our case, we have calculated fidelities of all the classes and shown them in Table 2. To obtain the experimental density matrices and fidelities we performed quantum state tomography of the outputs of all the circuits (see [43, 44] for

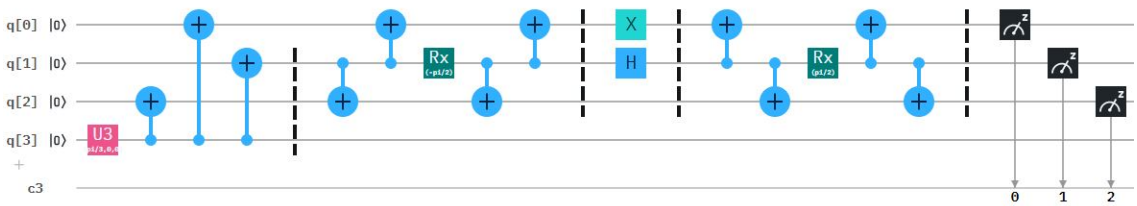
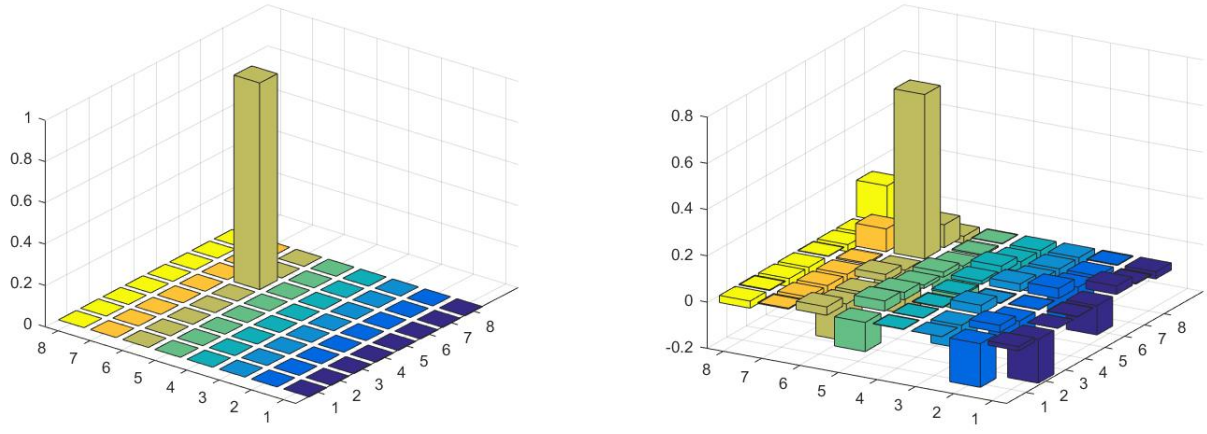


Figure 4: (Color online) Circuits for the simulation of noisy three-player quantum dilemma game for Class VIII as implemented on IBM Quantum Experience.



(a) (b)

Figure 5: (Color online) Graphical representation of (a) the theoretical and (b) the real part of the experimentally reconstructed density matrices for Class VII.

Action	Theoretical	Experimental (IBM)	Experimental (NMR)	Fidelity	Error in Payoff(%)	Processor
Class I	-3.75	-2.7689	-3.65	0.6753	26	vigo
Class II	-3.75	-2.751	-3.44	0.7813	26.64	ourense
Class III	-1.833	-0.7384	-1.59	0.844	59.71	ourense
Class IV	2	2.3275	1.26	0.9236	16.3	ourense
Class V	0	-0.0676	0.4	0.9346	6.76	essex
Class VI	-5.67	-4.608	-5.39	0.9317	18.73	ourense
Class VII	6.33	4.459	5.92	0.843	29.5	ibmqx2
Class VIII	6.33	4.03	6.34	0.8416	36.3	ibm_16_melbourne
Class IX	4.75	3.2	4.83	0.6753	32.6	ibmqx2
Class X	-1.833	-0.8098	-1.87	0.6314	55.8	vigo

Table 2: Results of the experiments performed using different quantum processors in IBM Q Experience for the payoff table parameters $p = 1, n = 9, q = 2$ considering state $x = 0$. Surprisingly, fidelity and error are not properly anti-correlated, i.e., there are instances where a class of higher fidelity than another, still yields a greater error in the payoff. An example here for different processors would be Class III and Class I. Another example, for the same processor, would be, Classes IV and VI. A possible explanation could be noise during readout of payoffs.

detail). The crux of the matter, is that we can reconstruct the three qubit density matrix of the output of the circuit using

$$\rho^E = \frac{1}{2^3} \sum_{i_1,i_2,i_3=0}^{3} T_{i_1 i_2 i_3} (\sigma_{i_1} \otimes \sigma_{i_2} \otimes \sigma_{i_3}), \tag{3}$$

where σ_j are respective Pauli matrices. The values of elements of the T matrix are obtained from the expectation values of these Pauli operators. For instance, in the case of Class VII, $U_7 = \sigma_x \otimes I \otimes \sigma_x$ is the strategy unitary, the experimentally reconstructed density matrix is shown with corresponding theoretical density matrix $\sigma = |101\rangle\langle 101|$ in Fig. 5. The fidelity of the output density matrices for all the classes are summarized in Table 2. Surprisingly, fidelity and error are not properly anti-correlated, i.e., there are instances where a class of higher fidelity than another, still yields a greater error in the payoff. Here it may be noted that the experiments are performed on different processors provided by IBM depending upon their availability.

4.2 Payoff table

Payoff of single player \$ is obtained by multiplying their respective payoffs from Table 1 with the probabilities obtained from the output of the experiment (as in [24]). The mean payoff per player $\langle \$ \rangle$ is defined as the numerical mean of the payoffs of each player $\langle \$ \rangle = \frac{\$1+\$2+\$3}{3}$.

It is also shown in the past that the payoffs for quantum Nash equilibrium deteriorate with noise [21]. However, in our case, assuming arbitrary values of parameters we obtained that quantum Nash equilibrium is $3\$_{\rm qu} = -4nx + 2n + p$, and classical Nash equilibrium for (q,q,q) is $\$_{\rm cl} = q\,(1 - x)$. As both quantum and classical Nash equilibrium values

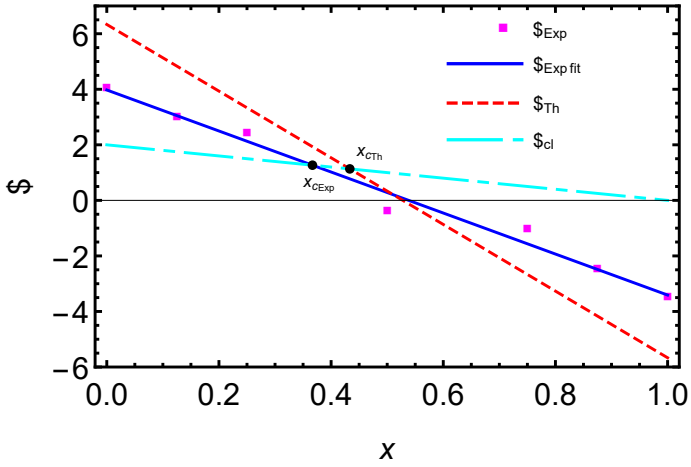
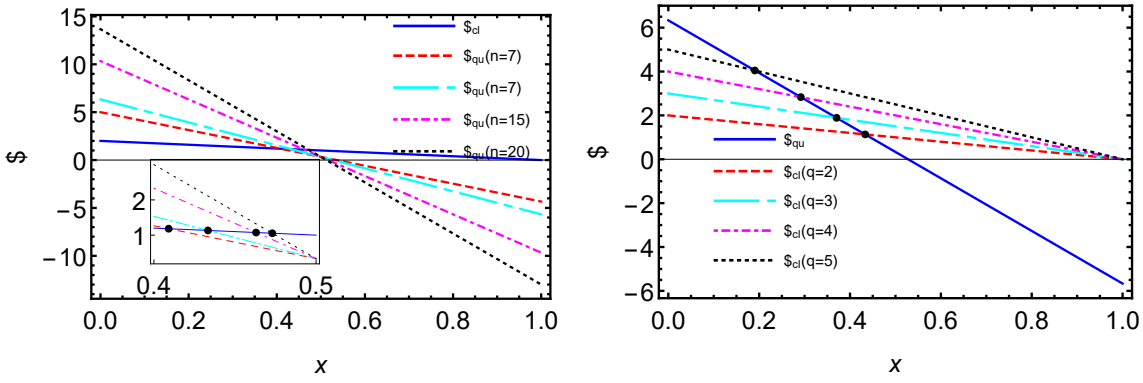


Figure 6: (Color online) Variation of payoffs in classical and quantum Nash equilibrium with corruption x. We have also shown the experimentally obtained payoff values. We assumed $p = 1$, $n = 9$, and $q = 2$. Here, $x_{c_{\rm Th}}$ and $x_{c_{\rm Exp}}$ represent critical values of corruption from theoretical and experimental results. Noise degrades quantum information, and thus not surprisingly, $x_{c_{\rm Th}} > x_{c_{\rm Exp}}$.



(a) (b)

Figure 7: (Color online) Variation of payoffs in classical and quantum strategies with corruption x and payoff parameters (a) n ($q = 2$) and (b) q ($n = 9$) with $p = 1$ for Class VIII considering rest of the parameters in $\{p, n, q\}$ as constant. The critical value of corruption x_c is marked by black dots in all cases.

decrease with corruption level x, the quantum advantage disappears after the point of intersection of these two curves. That intersection can be obtained as the critical value of corruption

$$x_c = \frac{2n + p - 3q}{4n - 3q}. \tag{4}$$

From Fig. 6, we obtained that experimentally $x_c = 0.363$ whereas the theoretical value is 0.428 [21], giving us an error of 15.18%. Note that the results obtained in [21] neglect decoherence after the initial state is prepared by the demon. However, here on top of that, gate errors in the implementation of the presently available SQUID based quantum computing facilities also play an important role in sabotaging the quantum advantage achievable in quantum games. Of course, a reduction in noise with improvement in technology will improve the outcome. Notice that for very high values of corruption, when classical strategy is a preferred choice, the experimental results show higher payoffs than theoretically expected in quantum Nash equilibrium. The experimental values of payoff can be improved using mitigated error method provided by IBM.

4.3 Variation in p, q, n

As we have discussed the general case of the game with arbitrary values of the individual payoff parameters, here we discuss the role of each of these parameters (assuming $0 < p < q < n$) on x_c. This would allow us to choose suitable payoff parameters if the noise level x is known, or if they cannot be varied, then to decide whether to employ the quantum or classical game for the problem in a practical situation.

We find that x_c increases (shifts to the right of the original value) if n is increased for a constant value of p and q. It implies that when the stakes of a game are high (large n), such that reward for winnings and the amount of losses are

very high, the quantum strategies are better in spite of corruption. Further increasing n saturates x_c to 0.5, which signifies that no matter what, if corruption is higher than 50% classical strategy will always outperform the quantum strategy. The results obtained are shown in Fig. 7 (a).

Note that the only quantum Nash equilibrium depends upon n, while classical Nash equilibrium is a function of q. Thus, an increase in q essentially leads to classical dominant strategy, i.e., classical strategy tends to be as efficient as the quantum strategy (cf. in Fig. 7 (b)). However, for large values of corruption there is no evident advantage as for maximum corruption classical Nash equilibrium is always zero.

These observation lead to conclusion that quantum systems are more prone to errors and deteriorate rapidly with an increase in the amount of corruption. Hence, errors in system may lead to loss of quantum advantage originally present as observed from the experimental value in Fig. 6 as well. Thus, in case of high errors, it is always better to stick to classical strategies from an outsider's perspective.

5 Conclusion

We have discussed a multiparty quantum game by generalizing the payoff table. Our result may find interesting applications in diverse fields, such as finance, social networks. Suppose a group of companies want to invest in a particular stock and have limited knowledge of the market statistics, then the financial situation of the stock simulates a three-person dilemma. In this case, if the stakes of the investments are high such that returns are great, but so are the losses, companies perform better if they use quantum strategies, provided the amount of preparation noise is less than 50%. Otherwise the classical strategy should be preferred. Similarly, in the situation that the classical dominant strategy equilibrium (q, q, q) has relatively higher payoff than previous cases, the present results may persuade companies to opt for classical strategies even for a small amount of source error.

We have performed an experiment for the noisy three-player quantum dilemma game and observed that the obtained results were less robust against noise than the corresponding results from NMR experiments. Further, it can be observed that due to additional errors (other than source error introduced in the noisy counterpart of the game) the advantage of quantum game over corresponding classical game disappears quickly. Similar studies for the generalizations of other games where quantum players perform better or the games where classical strategies are always preferable can be performed to study the role of various payoff parameters in those cases. The present experimental implementation of the noisy quantum game on a small noisy quantum computer establishes a practical quantum advantage in game theory. However, in view of noisy intermediate-scale quantum (NISQ) technology [45] around the corner, i.e., quantum computing infrastructure with 50-100 qubits, this advantage can be exploited for several applications, such as in quantum machine learning [46]. This can be further extended to the iterated version of the game where it is performed more than once, and rational players decide their strategies depending upon their opponents' previous decision. The results from the experiment performed on NMR was more accurate, showing that the NMR-based quantum computer is less noisy. To obtain a quantitative perception of that we performed quantum state tomography here, which shows that higher fidelity of experimentally generated state does not necessarily mean smaller errors, i.e., fidelity and errors are not properly anti-correlated.

In the end, we would like to stress on the recent studies connecting Bell nonlocality [47, 48], a quantum secure direct communication scheme [49], and security of quantum key distribution schemes [50] with game theory. In view of these works, in principle, all quantum cryptographic schemes (see [51, 52] for a review) can be viewed from the perspective of game theory, as a game to perform cryptanalysis and obtain security proofs. For example, measurement-device-independent and device-independent as well as entangled state based quantum key distribution schemes, such as Ekert's scheme [53], can be viewed as a three-party game involving Alice, Bob and Eve. A future work is planned to rigorously analyze the best strategy of Alice and Bob and that of Eve using a game theoretic approach. We hope the present results will be helpful in the application of quantum strategies in game theory, and in turn in their applications in quantum technologies in general, and quantum cryptography in particular.

Acknowledgement

AP and RS acknowledge the support from the QUEST scheme of Interdisciplinary Cyber Physical Systems (ICPS) programme of the Department of Science and Technology (DST), India, Grant No.: DST/ICPS/QuST/Theme-1/2019/14. KT acknowledges the financial support from the Operational Programme Research, Development and Education - European Regional Development Fund project no. CZ.02.1.01/0.0/0.0/16_019/0000754 of the Ministry of Education, Youth and Sports of the Czech Republic. Authors also thank Prof. Anil Kumar and Dr. Abhishek Shukla for their interest and technical comments on this work.

References

[1] Flitney, A. P., Abbott, D.: An introduction to quantum game theory. Fluctuation and Noise Letters **2**, R175–R187 (2002)

[2] Piotrowski, E. W., Sładkowski, J.: An invitation to quantum game theory. International Journal of Theoretical Physics **42**, 1089–1099 (2003)

[3] Li, X., Gao, L., Li, W.: Application of game theory based hybrid algorithm for multi-objective integrated process planning and scheduling. Expert Systems with Applications **39**, 288–297 (2012)

[4] Elhenawy, M., Elbery, A. A., Hassan, A. A., Rakha, H. A.: An intersection game-theory-based traffic control algorithm in a connected vehicle environment. in: 2015 IEEE 18th international conference on intelligent transportation systems pp. 343–347 IEEE (2015)

[5] Laszka, A., Szeszlér, D., Buttyán, L.: Game-theoretic robustness of many-to-one networks. in: International Conference on Game Theory for Networks pp. 88–98 Springer (2012)

[6] Dodis, Y., Rabin, T., et al.: Cryptography and game theory. Algorithmic game theory pp. 181–207 (2007)

[7] Li, A., Yong, X.: Entanglement guarantees emergence of cooperation in quantum prisoner's dilemma games on networks. Scientific reports **4**, 1–7 (2014)

[8] Nawaz, A., Toor, A. H.: Dilemma and quantum battle of sexes. Journal of Physics A: Mathematical and General **37**, 4437 (2004)

[9] Meyer, D. A.: Quantum strategies. Physical Review Letters **82**, 1052 (1999)

[10] Eisert, J., Wilkens, M., Lewenstein, M.: Quantum games and quantum strategies. Physical Review Letters **83**, 3077 (1999)

[11] Du, J., Xu, X., Li, H., Zhou, X., Han, R.: Playing prisoner's dilemma with quantum rules. Fluctuation and Noise Letters **2**, R189–R203 (2002)

[12] Werner, R. F.: Optimal cloning of pure states. Physical Review A **58**, 1827 (1998)

[13] Tomamichel, M., Fehr, S., Kaniewski, J., Wehner, S.: A monogamy-of-entanglement game with applications to device-independent quantum cryptography. New Journal of Physics **15**, 103002 (2013)

[14] Fritz, T., Sainz, A. B., Augusiak, R., et al.: Local orthogonality as a multipartite principle for quantum correlations. Nature communications **4**, 1–7 (2013)

[15] Oppenheim, J. and Wehner, S., 2010. The uncertainty principle determines the nonlocality of quantum mechanics. Science, 330(6007), pp.1072-1074.

[16] Anshu, A., Høyer, P., Mhalla, M., Perdrix, S.: Contextuality in multipartite pseudo-telepathy graph games. Journal of Computer and System Sciences **107**, 156–165 (2020)

[17] Popescu, S.: Nonlocality beyond quantum mechanics. Nature Physics **10**, 264–270 (2014)

[18] Chen, C., Dong, D., Dong, Y., Shi, Q.: A quantum reinforcement learning method for repeated game theory. in: 2006 International Conference on Computational Intelligence and Security volume 1 pp. 68–72 IEEE (2006)

[19] Clausen, J., Briegel, H. J.: Quantum machine learning with glow for episodic tasks and decision games. Physical Review A **97**, 022303 (2018)

[20] Benjamin, S. C., Hayden, P. M.: Multiplayer quantum games. Physical Review A **64**, 030301 (2001)

[21] Johnson, N. F.: Playing a quantum game with a corrupted source. Physical Review A **63**, 020302 (2001)

[22] Daskalakis, C., Papadimitriou, C. H.: Three-player games are hard. in: Electronic colloquium on computational complexity volume 139 pp. 81–87 (2005)

[23] Daskalakis, C., Goldberg, P. W., Papadimitriou, C. H.: The complexity of computing a Nash equilibrium. SIAM Journal on Computing **39**, 195–259 (2009)

[24] Mitra, A., Sivapriya, K., Kumar, A.: Experimental implementation of a three qubit quantum game with corrupt source using nuclear magnetic resonance quantum information processor. Journal of Magnetic Resonance **187**, 306–313 (2007)

[25] Kolenderski, P., Sinha, U., Youning, L., et al.: Aharon-Vaidman quantum game with a Young-type photonic qutrit. Physical Review A **86**, 012321 (2012)

[26] Pinheiro, A., Souza, C., Caetano, D., et al.: Vector vortex implementation of a quantum game. JOSA B **30**, 3210–3214 (2013)

[27] Schmid, C., Flitney, A. P., Wieczorek, W., et al.: Experimental implementation of a four-player quantum game. New Journal of Physics **12**, 063031 (2010)

[28] Zhou, L., Kuang, L.-M.: Proposal for optically realizing a quantum game. Physics Letters A **315**, 426–430 (2003)

[29] Solmeyer, N., Linke, N. M., Figgatt, C., et al.: Demonstration of a Bayesian quantum game on an ion-trap quantum computer. Quantum Science and Technology **3**, 045002 (2018)

[30] Glazer, J., Ma, C.-t. A.: Efficient allocation of a "prize"-King Solomon's dilemma. Games and Economic Behavior **1**, 222–233 (1989)

[31] Teng, Y., Jones, R., Marusich, L., et al.: Trust and situation awareness in a 3-player diner's dilemma game. in: 2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA) pp. 9–15 IEEE (2013)

[32] Bankes, S.: Exploring the foundations of artificial societies: Experiments in evolving solutions to iterated N-player prisoner's dilemma. in: Artificial Life IV pp. 337–342 MIT Press Cambridge, MA (1994)

[33] Morgan, J. P., Chaganty, N. R., Dahiya, R. C., Doviak, M. J.: Let's make a deal: The player's dilemma. The American Statistician **45**, 284–287 (1991)

[34] Özdemir, Ş. K., Shimamura, J., Imoto, N.: Quantum advantage does not survive in the presence of a corrupt source: optimal strategies in simultaneous move games. Physics Letters A **325**, 104–111 (2004)

[35] Chen, J.-L., Kwek, L. C., Oh, C. H.: Noisy quantum game. Physical Review A **65**, 052320 (2002)

[36] Chen, L., Ang, H., Kiang, D., Kwek, L., Lo, C.: Quantum prisoner dilemma under decoherence. Physics Letters A **316**, 317–323 (2003)

[37] IBM Q. https://www.research.ibm.com/ibm-q/. Accessed on September 2019

[38] Sisodia, M., Shukla, A., Thapliyal, K., Pathak, A.: Design and experimental realization of an optimal scheme for teleportation of an n-qubit quantum state. Quantum Information Processing **16**, 292 (2017)

[39] Sisodia, M., Shukla, A., de Almeida, A. A., Dueck, G. W., Pathak, A.: Circuit optimization for IBM processors: A way to get higher fidelity and higher values of nonclassicality witnesses. arXiv preprint arXiv:1812.11602 (2018)

[40] Tian, J., Uchida, N.: Monkeys in a prisoner's dilemma. Cell **160**, 1046–1048 (2015)

[41] Shukla, A., Sisodia, M., Pathak, A.: Complete characterization of the directly implementable quantum gates used in the IBM quantum processors. arXiv preprint arXiv:1805.07185 (2018)

[42] Sousa, P. B., Ramos, R. V.: Universal quantum circuit for n-qubit quantum gate: A programmable quantum gate. arXiv preprint quant-ph/0602174 (2006)

[43] Sisodia, M., Shukla, A., Pathak, A.: Experimental realization of nondestructive discrimination of Bell states using a five-qubit quantum computer. Physics Letters A **381**, 3860–3874 (2017)

[44] Vishnu, P., Joy, D., Behera, B. K., Panigrahi, P. K.: Experimental demonstration of non-local controlled-unitary quantum gates using a five-qubit quantum computer. Quantum Information Processing **17**, 274 (2018)

[45] Preskill, J.: Quantum computing in the NISQ era and beyond. Quantum **2**, 79 (2018)

[46] Torlai, G., Melko, R.: Machine-learning quantum states in the NISQ era. Annual Review of Condensed Matter Physics **11** (2019)

[47] Brunner, N., Linden, N.: Connection between Bell nonlocality and Bayesian game theory. Nature communications **4**, 2057 (2013)

[48] Iqbal, A., Chappell, J. M., Abbott, D.: The equivalence of Bell's inequality and the Nash inequality in a quantum game-theoretic setting. Physics Letters A **382**, 2908–2913 (2018)

[49] Kaur, H., Kumar, A.: Game-theoretic perspective of ping-pong protocol. Physica A: Statistical Mechanics and its Applications **490**, 1415–1422 (2018)

[50] Krawec, W. O., Miao, F.: Game theoretic security framework for quantum key distribution. in: International Conference on Decision and Game Theory for Security pp. 38–58 Springer (2018)

[51] Pathak, A.: Elements of quantum computation and quantum communication. Taylor & Francis, New York (2013)

[52] Thapliyal, K., Pathak, A., Banerjee, S.: Quantum cryptography over non-Markovian channels. Quantum Information Processing **16**, 115 (2017)

[53] Ekert, A. K.: Quantum cryptography based on Bell's theorem. Phys. Rev. Lett. **67**, 661 (1991)

Appendix: Reconstructed density matrix

The real and imaginary parts of the experimentally obtained density matrix by performing quantum state tomography are

$$\mathrm{Re}(\rho^E) = \begin{bmatrix} 0.018 & -0.188 & -0.001 & 0.002 & -0.126 & 0.036 & -0.004 & -0.031 \\ -0.188 & 0.030 & -0.100 & 0.002 & 0.044 & -0.160 & -0.019 & -0.003 \\ -0.001 & -0.100 & 0.037 & -0.003 & 0.033 & -0.056 & -0.049 & -0.035 \\ 0.002 & 0.002 & -0.003 & 0.007 & 0.015 & 0.030 & -0.042 & -0.039 \\ -0.126 & 0.044 & 0.033 & 0.015 & 0.022 & -0.189 & -0.006 & -0.002 \\ 0.036 & -0.160 & -0.056 & 0.030 & -0.189 & 0.711 & 0.099 & -0.030 \\ -0.004 & -0.019 & -0.049 & -0.042 & -0.006 & 0.099 & 0.018 & -0.022 \\ -0.031 & -0.003 & -0.035 & -0.039 & -0.002 & -0.030 & -0.022 & 0.157 \end{bmatrix}$$

and

$$\mathrm{Im}(\rho^E) = \begin{bmatrix} 0 & 0.229 & 0.002 & 0.009 & 0.142 & -0.042 & -0.028 & -0.009 \\ -0.229 & 0 & 0 & -0.001 & -0.001 & 0.148 & -0.010 & -0.034 \\ -0.002 & 0 & 0 & 0.002 & -0.011 & 0.014 & 0.053 & 0.010 \\ -0.009 & 0.001 & -0.002 & 0 & 0.014 & 0.017 & 0.004 & 0.030 \\ -0.142 & 0.001 & 0.011 & -0.014 & 0 & 0.158 & 0.004 & -1.108 \\ 0.042 & -0.148 & -0.014 & -0.017 & -0.158 & 0 & -1.103 & 0.058 \\ 0.028 & 0.010 & -0.053 & -0.004 & -0.004 & 1.103 & 0 & 0.024 \\ 0.009 & 0.034 & -0.010 & -0.030 & 1.108 & -0.058 & -0.024 & 0 \end{bmatrix},$$

respectively, while theoretical density matrix in the corresponding case is given by $\sigma = |\psi\rangle\langle\psi| = |101\rangle\langle 101|$, where $|\psi\rangle = J^{\dagger} \cdot U_7 \cdot J\,|000\rangle$.

Office Spirituality and Job pleasure: Exploring the Mediating Effect of Office Citizenship Conduct

Uday Sai
Student
University School of Management and Entrepreneurship
Delhi Technological University (formerly Called as Delhi College of Engineering), Delhi, India

Abstract : This decade witness converted interest in explore studies on workplace spirituality. The present research investigates the relationship between workplace spirituality and job satisfaction with the help of two approaches. First, with the help of necessary condition analysis (NCA), it is exploring whether workplace spirituality is a required condition for the occurrence of Job satisfaction. Second, the mediation effect of organizational citizenship behavior (OCB) in the relationship amid WSP and JS is examined using Baron and Kenny's (1986) technique of mediation analysis. A sample of 194 managers working in insurance companies is collected for the study. The study reports a positive correlation between all three variables of the study. Further, WSP does not come out as a necessary condition for JS. A partial mediation effect of OCB is reported. Practical implications and limitations are also discussed in the article.

Index Terms - Workplace Spirituality, Organizational Citizenship Behavior, Job satisfaction, Necessary Condition Analysis

I. INTRODUCTION:-

Office spirituality (OSP) has started to get converted concentration from study scholars and practitioners both (Giacalone, Jurkiewicz, & Fry, 2005). Way back in 1980, Toffler (1980) stated 'technology' as a third wave of organizations. Nowadays, WSP has increasingly resorted to a revolutionary antidote for contemporary human resource-related issues. Theoretical and experimental studies on the definition, notion, construct and effectiveness of WSP have expanded considerably in recent times (Houghton, Neck, & Krishnakumar, 2016). WSP reduces organizational stress and improves decision-making (Tischler et al., 2002), highly satisfied workforce at the workplace (Harung et al., 1996), effective and efficient performance at workplace (Duchon & Plowman, 2005, Malik et al., 2011; Malieki et al., 2012), a higher level of job involvement and greater compliance with ethical constructs such as honesty and trust (Krishnakumar & Neck, 2002), and lower levels of psychological stress (Adams et al., 2003). Spirituality at the workplace is not a fringe idea (Petchsawanga & Duchon, 2012), rather it is concerned with human's deep psychological needs such as holistic development of the personality and expansion of human virtues such as compassion, trust, and faith. At a broader level, WSP inspires an employee to start searching for real essence and aim of human life (Gull & Doh, 2004). Moore and Casper (2006) categorized the benefits of WSP in three broad levels—individual level, organizational level, and societal level. At a societal level, WSP enhances people's trust in goodness (Miller, 2001). At an organizational level, the profitability and productivity of the organizations increase with the implementation of WSP. Organizations soaked in WSP shows greater work efficiency and a higher rate of return (Jurkiewicz & Giacalone, 2004). Researchers including Neck and Milliman (1994); Pandey, Gupta, and Arora (2009); Malik et al. (2011), Garg (2017a) and Malieki et al. (2012) supported positive linkage between WSP and performance of the organization. Further, at an individual level, employees' physical and mental health is positively associated with WSP (Jurkiewicz & Giacalone, 2004; Krahnke et al., 2003). WSP is reported to have a positive effect on the confidence level of the employees (Krahnke et al., 2003). Organizational commitment of the employee increases with the experience of WSP (Garg, 2017b; Rego & Cunha, 2008), emotional quotient of employees (Hartsfield, 2003), employees' motivation level (Bhunia & Das, 2012), commitment towards career (De Klerk, Boshoff, & Van Wyk 2006), ethical and cultural values (Issa & Pick, 2011) and self-efficacy (Garg, 2014). Robert et al. (2006) reported a positive connection between WSP and Job Satisfaction (JS). Various other studies also indicated linkage WSP and JS but none of the previous studies have explored the connection using necessary condition analysis (NCA). The article is roughly categorized into four major segments. The first segment of the article provides diverse but enriching definitions and viewpoints of different research scholars around the globe. Indian dimensions/ constructs of WSP are discussed in the second part of the article. Third, the necessity of independent variable (WSP) for JS is explored using a new and innovative statistical tool—NCA. The mediating effect of organizational citizenship behavior (OCB) is examined using method devised and used by Baron and Kenny (1986). The theoretical and practical significance of the findings along with propositions for future researchers is elaborated in the last part of the article.

II. LITERATURE REVIEW

Workplace Spirituality—Definitions and Concept

The available literature on WSP is full of different definitions of WSP. Kinjerski and Skrypnek (2004) reported as many as 70 definitions of WSP but unfortunately, none of the definitions got wide and unanimous acceptance and recognition from all academicians. Few definitions of WSP are tabulated as follows: WSP has been differentiated from religion. It is not concerned with converting employees into a religious or sect-based belief system (Cavanagh, 1999). WSP is based on individual choices of values and work perception of the employees. It is about individuals who perceive themselves as spiritual creatures, whose inner selves demand nourishment at the workplace, who enjoy self-identification and purpose of life through work experience and ones who have developed a sense of connectedness with others (Ashmos & Duchon, 2000; Milliman Czaplewski, & Ferguson, 2003). Shrimad Bhagwad Gita defines WSP in terms of excellence in action at the workplace. It explains 'yoga karmas' which means 'excellence in action is yoga' (Shrimad Bhagwad Gita, verse 2.50). Researchers have provided diverse nomenclature to WSP: 'Experience of inner self' (Dillard, 1982), 'superintelligence' (Zohar &Marshall, 2000), 'common purpose' (Kinjerski & Skrypnek, 2004), 'inclusiveness and interconnectedness' (Marcqus et al., 2005), 'inner consciousness' (Guillory, 2000), 'feelings that energize action' (Dehler & Welsh, 1994), 'sacred force that implies life' (Nash &

Mclennan, 2001), 'compassion' (McCormick, 1994), etc. Although there is no agreement of researchers on the universally accepted definition of WSP, dimensions of WSP are agreed in principle.

Workplace Spirituality and Job Satisfaction

Porter and Lawler (1968) defined JS as 'a unidimensional construct; that is, JS is people's affective (emotional) response to their current job conditions'. JS is pivotal to HR managers because it is largely known for triggering attrition (Sturges & Guest, 2001). Jernigan et al. (2002) define JS as a positive attitude of employees for work, work environment, colleagues and also for larger organizational context. The broader context of JS includes satisfaction with salary, job responsibility, peers and juniors and with company's policies and procedures. However, a narrowly defined concept of JS includes only constructive and positive psychological assessment of job profile (Locke & Latham, 1990). In summation, it is a job attitude that is largely determined by the employee's experience at the workplace. This experience is shaped by the physical work environment, the relationship of colleagues among themselves, leadership style and also by the degree of decentralization and power delegation (Kersley et al., 2006). De Klerk et al. (2006) highlighted the fact that work, organization, and working environment are increasingly acquiring center stage in one's life. People spend more time at the workplace or informal work-related activities than any other personal stuff. Thus, employees aspire to fulfill their spiritual needs at the workplace itself (Garg, 2018). Institutionalization of WSP ensures meaningful and purposeful work along with social affiliation and thus it leads to spiritual fulfillment (Der Walt & De Klerk, 2014). Spiritually fulfilled employees have higher levels of satisfaction. WSP is increasingly being lauded as a promoter of JS. WSP leads to desired outcomes such as worker motivation and JS (Krishnakumar & Neck, 2002). Pawar (2009) and Milliman et al. (2003) also reported a constructive linkage between two variables (WSP and JS). Other researchers who confirmed positive association includes Chawla and Guda (2010), Kolodinsky et al. (2008), Usman and Danush (2010) and many more. The present study does not only explore the correlation between WSP and JS, but it also aspires to investigate the necessity of WSP for JS with the help of a novel but an effective statistical tool —NCA.

Workplace Spirituality and Organizational Citizenship Behaviour

According to Bateman and Organ (1983), OCB is 'individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and in the aggregate promotes the effective functioning of the organization'. It is a multi-dimensional construct that is commonly studied with the help of the following five dimensions (Garg & Sharma, 2015). 1. Courtesy: Proactive gestures that consider consulting with other workers in the organization before acting, giving advance notice and passing along information. 2. Altruism: Voluntary behavior of the staff member through which an employee tries to aid colleagues in a particular task or work assignment. 3. Conscientiousness: Employee performance his or her assigned tasks according to their role expectations. 4. Civic virtue: Engrossment which the employee illustrates in the political life of the organization. 5. Sportsmanship: Abstaining from complaining about trivial matters. WSP was reported to be a strong predictor of OCB (Fatahi, 2007). A statistically significant relationship between WSP and OCB was observed by Farhangi et al. (2006). In a famous experimental study, Geh (2010) theorized and examined a hypothetical model based on the association of WSP and OCB. Like other researchers such as Harmer (2008) and Liu (2008), the author also concluded a positive association between two variables (WSP and OCB). In summation, the above-quoted piece of theoretical and empirical researches spelled out the potential of WSP in dealing with the challenges and problems of contemporary organizations. However, unfortunately, the mechanisms through which WSP impacts JS have not been explored extensively. The present study tends to plug the stated research gap by investigating the mediation effect of OCB. The proposition of the study is: OCB mediates the association of WSP and JS. Secondly, the study explores the necessity of WSP for JS using NCA. Accordingly, the second proposition is: WSP is an indispensable condition for JS.

III. METHODOLOGY

NCA has been used to identify compulsory or essential conditions. The novel statistical technique is utilized to investigate whether WSP is an indispensable (necessary) requirement for JS. A necessary cause is a constraint, a barrier, an obstacle, a bottleneck that must be managed to allow a desired outcome to exist. Each single necessary cause must be in place, as there is no additive causality that can compensate for the absence of the necessary cause. (Dul, 2016, pp. 11)

Fulfillment of necessary does not guarantee the outcome but non-existence of necessary condition leads to guaranteed failure. Outcome cannot exist without the occurrence of necessary condition. Dul (2016), father of NCA, further elaborated that every necessary condition is indispensable. Further, the mediation effect of OCB is investigated using Baron and Kenny's (1986) approach which confirms the mediation effect if the following conditions are satisfied:-

1. Independent/predictor variable (WSP) is significantly related to criteria/dependent variable (JS).
2. The mediating variable (OCB) is reported to have a significant relationship with the dependent variable.
3. Predictor variable has a statistically significant relationship with the mediator.
4. Keeping mediating variable controlled, if the degree of association between predictor and criteria variable condensed significantly then partial mediation is concluded. And if the relationship between independent and independent variable ceases, then full mediation in confirmed.

IV. RESEARCH METHODOLOGY.

Research Framework Data Collection and Sample : The study used descriptive research design. A sample of 194 junior- and middle-level managers working in insurance companies was collected using convenience sampling. Respondents were approached through an online survey using a structured questionnaire comprising of measures of WSP, job performance, JS and OCB. The data was collected from several life and non-life insurance companies such as LIC, HDFC Ergo, GIC Re and United India Insurance Company Limited. Sampling captured nearly equal proportion of male (51%) and female (49%) respondents. As many as 25 respondents were below the age of 25 years, 88 employees were of age ranging from 26 to 35 years. Fifty employees were of the age group of 36 to 45 years and the rest of the employees (31 employees) were above the age of 45 years. The average age of the participants was 38.6 years. Also, as high as 71 percent of respondents were graduates while 19 percent were diploma holders and 10 percent was postgraduate. Figure 1 defines the direct relationship between workplace spirituality and job satisfaction. Figure 2 describes the mediation model wherein OCB is proposed to mediate the relationship between WSP and job satisfaction



Figure I: - Direct Path
Source:-From previous studies.

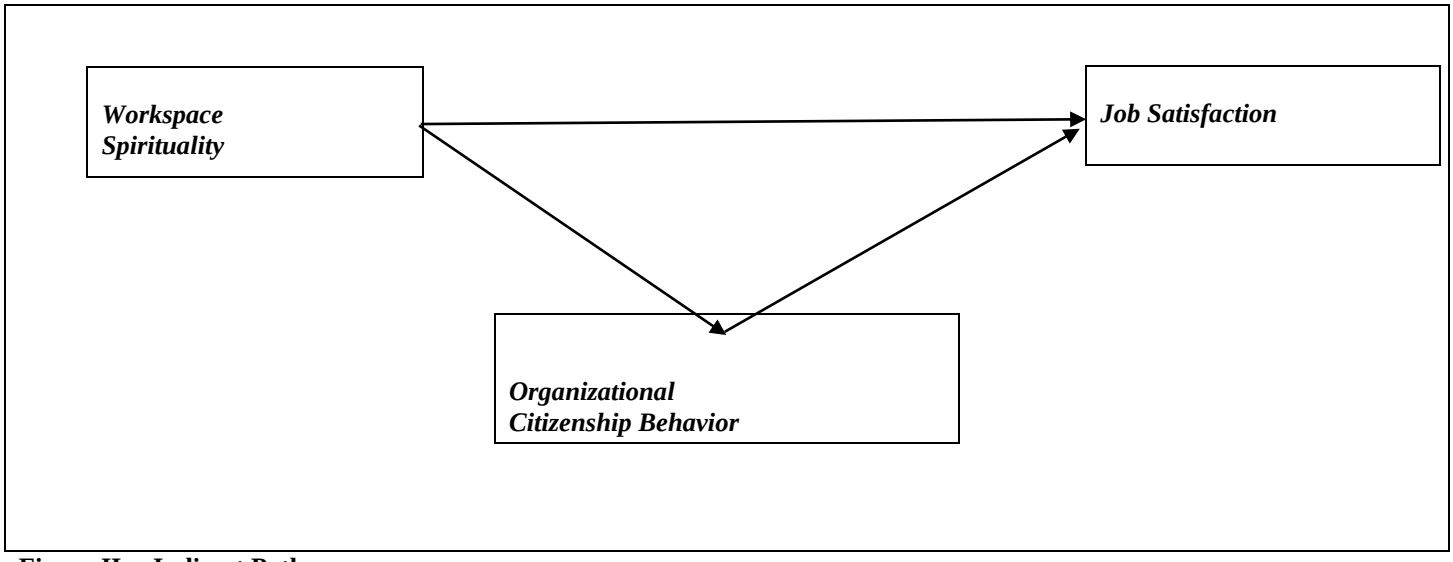


Figure II:- Indirect Path
Source:- From previous studies.

V. RESULTS

It is reported that all constructs of WSP (swadharma, lokasangraha, sense of community, authenticity, karma capital and kritagayata) and overall WSP have a statistical considerable affirmative correlation with five constructs of OCB and also with overall OCB. Also, encouraging positive association is also observed between WSP and JS ($r = 0.75$). Since both variables are reported to move in the same fashion then further analysis with the help of regression and NCA is feasible. Results are complementary with the results of other previous studies conducted in different cultural settings and with the help of different dimensions of WSP. Similar results are replicated with Indian constructs of WSP too. The current study does not end here, however the relationship between WSP and JS is investigated further using NCA. For the simplification of analysis, NCA is divided into two simple and basic steps considering WSP as independent variable and JS as a dependent variable.

1. A scatter diagram is plotted considering independent variable on X-axis and dependent variable on Y-axis. Subsequently, NCA is designed to produce a ceiling line that differentiates the plot in two different zones. Zone 1 is called an empty zone and this zone does not have any observation. Zone 2 is full zone having nearly all observations of the study. 2. Now, quantitative and statistical parameters are calculated to determine accuracy of plotted ceiling line. Accuracy has been defined mathematically as number of observations in full zone divided by total observations. Accuracy equal to or above 95 per cent is desirable (Dul, 2016). Subsequently, effect size, which is an amplitude of necessity, is calculated. Value of effect size lies between 0 and 1 and could be analysed in the following way:

$0 < d < 0.1$ represents small effect size, $0.1 \leqslant d < 0.3$ represents medium effect, $0.3 \leqslant d < 0.5$ concludes large effect and $0.5 \leqslant d < 1.0$ confirms very large effect size.

According to Dul (2016), empty upper-left corner region of the graph indicates the necessity of the independent variable. Two types of statistics are produced by NCAceiling envelopment with free disposal hull (CE-FDH) and ceiling regression with free disposal hull (CE-

RDH). Since CE-FDH produces a better result (Dul, 2016), thus it has been preferred for analysis purpose. Threshold limit of 0.1 is essential to draw conclusive evidence of the existence of necessity of predictor variable (Dul, 2016).

Baron and Kenny's (1986) methodology of mediation analysis has four steps. Table 3 depicts the results of four sequential steps labelled as four different models. The first condition of mediation analysis is the presence of a significant relationship between an independent variable and dependent variable. Accordingly, overall WSP was regressed on JS (dependent variable) in Model 1, and it was significant and positive ($\beta = 0.75$, $p < 0.05$), which confirms the fulfilment of the first condition. Condition 2 requires a significant relationship between an independent variable (here it is WSP) and proposed mediator variable (OCB) and interestingly β value of 0.85 at $p < 0.05$ also confirms the same. Condition 3 demands a significant relationship between mediator variable and dependent variable. This condition is also met with β value of 0.73 at $p < 0.01$. Condition 4 analyses the relationship between IV and DV on controlling the effect of mediator variable. Here, hierarchical regression was used to control OCB and it was reported that the β value decreases from 0.73 to 0.70 on controlling OCB. Since a correlation coefficient of WSP and JS reduced on controlling OCB but still, it remained significant thus partial mediation is confirmed.

VI. CONCLUSION

Earlier scholars have decided a positive association between OSP and JP but mediation studies in the domain of WSP are scared. The present article explores the probable mediation effect of OCB. The mediating variable illustrates the mechanism through which the effect of an independent variable transmits to the dependent variable. According to results, WSP affects OCB and in turn, OCB affects JS. All six dimensions of WSP have been extracted from ancient Indian scriptures and also from previous academic and empirical research explorations in the Indian context. In nutshell, the present study in Indianization of spirituality studies. Another vital contribution of the study is the application of NCA in the study of WSP. One of the propositions is proved correct and another proposition is proved wrong. The article concludes that WSP is not a necessary condition for JS. Further, it is also reported that OCB mediates partially the relationship between WSP and JS. In nutshell, the results suggest that those employees who experience spirituality at the workplace will indulge in OCB and this inclination towards OCB will lead to a higher level of JS among them. Several practical implications could be inferred from the present study. As spirituality is highly correlated with JS, thus employers may try to incorporate spiritual intent and values at the workplace. This spiritual intent must manifest itself in the form of more enlightened and enriching organizational vision, missions and practices. Fulfillment of the spiritual needs of the employees may lead to a more collaborative and responsible workforce. This constructive interaction and collaboration form the basis of highly satisfied workforce. Organizations around the globe spend millions on increasing JS at workplaces and if some focus is given to the fundamental constructs of WSP, then there will be much more to be gained. As usual, the present study has its limitations too. First, owing to the collection of information from the same source and also at the same time, there is a risk of a common method error. Future studies in the area may apply double source technique or longitudinal approach to mitigate the issue. Second, the sample of the present is derived from insurance companies thus results could be generalized for the insurance industry but generalization for other industries requires further inquiries and explorations. Future researchers may be required to adopt both qualitative and quantitative research approaches. Exploration of the probable relationship of WSP and employee level outcomes including employee engagement, organizational commitment, etc., could help in crystallizing the domain of WSP. The probable mediating effect of other intervening variables including emotions, trust, perceived organizational support, etc., could enrich knowledge and understanding of the field. Further, studies on the moderation effect are scarce in the domain of WSP. One could contribute by investigating the moderating effect of variables such as gender, experience, and age.

VII. Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, and for publication of this piece of writing.

VIII. Funding

The author received no economic support for the research, and for publication of this piece of writing

XI. Acknowledgment

I would like to show my warm thanks to my parents who supported me at every bit and without whom it was impossible to accomplish the end task. The physical support and the technical contribution of "*Delhi Technological University* – Faculty, Without their support and care, it was impossible to reach the goal. Last but not the least; emotional support can also be considered as a key to success. So, I wish to thank my parents, sister, and my friends who supported me throughout my research.

X. REFERENCES

[1] Harung, H.S., Heaton, D.P., Graff, W.W., & Alexander, C.N. (1996). Peak performance and higher states of consciousness: A study of world-class performers. *Journal of Managerial Psychology*, 11(4), 3–23.
[2] Milliman, J., Czaplewski, A. J., & Ferguson, J. (2003). Workplace spirituality and employee work attitudes: An exploratory empirical assessment. *Journal of Organizational Change Management,* 16, 426–447.
[3] Pandey, A., Gupta, R. K., & Kumar, P. (2016). Spiritual climate and its impact on learning in teams in business organizations. *Global Business Review*, 17(3 Suppl.), 1S–14S.
[4] Rego, A., & Cunha, M. P. (2008). Workplace spirituality and organizational commitment: An empirical study. *Journal of Organizational Change Management,* 21(1), 53–70.
[5] Singh, K. (2004). Impact of HR practices on perceived firm performance in India, *Asia Pacific. Journal of Human Resources*, 42(3), 301–317.
[6] Truss, C., Shantz, A., Soane, E., Alfes, K., & Delbrige, R. (2013). Employee engagement, organizational performance and individual well-being: Evidence, developing the theory. *The International Journal of Human Resource Management*, 24, 2657–2669.
[7] Turner, J. (1999). Spirituality in the workplace. CA *Magazine*, 132(10), 41–42.
[8] Usman, A., & Danish, R.Q. (2010). Spiritual consciousness in banking managers and its impact on job satisfaction. *International Business Research*, 3(2), 65–72.

applied sciences

MDPI

Review

Rainfall Induced Landslide Studies in Indian Himalayan Region: A Critical Review

Abhirup Dikshit [1], Raju Sarkar [2], Biswajeet Pradhan [1,3,*], Samuele Segoni [4] and Abdullah M. Alamri [5]

1 Centre for Advanced Modeling and Geospatial Information Systems (CAMGIS), University of Technology, Sydney, NSW 2007, Australia; abhirup.dikshit@student.uts.edu.au
2 Department of Civil Engineering, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India; rajusarkar@dce.ac.in
3 Department of Energy and Mineral Resources Engineering, Sejong University, Choongmu-gwan, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea
4 Department of Earth Sciences, University of Firenze, Via Giorgio La Pira, 4, 50121 Florence, Italy; samuele.segoni@unifi.it
5 Department of Geology & Geophysics, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; amsamri@ksu.edu.sa
* Correspondence: Biswajeet.pradhan@uts.edu.au

Received: 25 February 2020; Accepted: 31 March 2020; Published: 3 April 2020



Abstract: Landslides are one of the most devastating and recurring natural disasters and have affected several mountainous regions across the globe. The Indian Himalayan region is no exception to landslide incidences affecting key economic sectors such as transportation and agriculture and often leading to loss of lives. As reflected in the global landslide dataset, most of the landslides in this region are rainfall triggered. The region is prone to 15% of the global rainfall-induced landslides, and thereby a review of the studies in the region is inevitable. The high exposure to landslide risk has made the Indian Himalayas receive growing attention by the landslides community. A review of landslides studies conducted in this region is therefore important to provide a general picture of the state-of-the-art, a reference point for researchers and practitioners working in this region for the first time, and a summary of the improvements most urgently needed to better address landslide hazard research and management. This article focuses on various studies ranging from forecasting and monitoring to hazard and susceptibility analysis. The various factors used to analyze landslide are also studied for various landslide zones in the region. The analysis reveals that there are several avenues where significant research work is needed such as the inclusion of climate change factors or the acquisition of basic data of highest quality to be used as input data for computational models. In addition, the review reveals that, despite the entire region being highly landslide prone, most of the studies have focused on few regions and large areas have been neglected. The aim of the review is to provide a reference for stakeholders and researchers who are currently or looking to work in the Indian Himalayas, to highlight the shortcomings and the points of strength of the research being conducted, and to provide a contribution in addressing the future developments most urgently needed to obtain a consistent advance in landslide risk reduction of the area.

Keywords: Indian Himalayas; landslides; GIS; remote sensing

1. Introduction

Landslides are the most frequent naturally occurring hazards that affect people and their livelihood worldwide. The frequency of occurrence in the Himalayan context is very large when compared with global events [1]. This review paper is an attempt to understand the research being undertaken to

understand, assess, and mitigate landslide scenarios in the Indian Himalayan region. The need for such a review was raised from the compilation of the global landslide disaster database by Froude and Petley, [1]. In their database, a total of 5318 non-seismic landslides occurred from 2004 to 2017, of which 3285 landslides were triggered by rainfall. In the context of the Indian Himalayas, during the same period, 580 landslides occurred with 477 triggered by rainfall, thereby contributing 14.52% of the global landslides. These number could be even higher; for instance, based on NASA GLC data, the number of landslides in the Indian Himalaya during 2007–2015 is 691 with 6306 casualties (https://data.nasa.gov/Earth-Science/Global-Landslide-Catalog/h9d8-neg4).

This region covers more than 12% of India's landmass and is very prone to landslides due to the fragile lithology, the complex geological setting, the high energy of the relief with steep slopes, and the high topographic roughness. Moreover, most of the area is seismically active and subject to extreme precipitations, and the situation has been further worsened with the increase in anthropogenic activities and the advent of climate change. Since it is a well-established fact that most of the landslides in this area have been primarily triggered by rainfall [2–4], the focus of the present review is only on the studies considering rainfall triggered landslides. Moreover, one of the main practical purposes of this review is to serve as a starting point for future projects which consider implementing territorial landslide early warning systems, and rainfall triggering landslides are the only ones that at present can be forecasted with a certain confidence. Studies on earthquake-induced landslides, snow avalanches, and Glacier Lake Outburst Flood (GLOF) were not considered in this review. This work also covers the various methodologies being adopted for landslide monitoring and analysis, as well as reports on various mitigation measures being undertaken along with the role of Geographic Information Systems (GIS), remote sensing, and the recent use of computational techniques.

2. Materials and Methods

2.1. Study Region

The Indian Himalayan region is one of the most diverse and heterogeneous area in terms of geology, lithology, rainfall distribution, land use/land cover, soil properties, and road and stream networks, which makes it highly prone to landslides. The region covers 16.2% (~500,000 km^2) of India's landmass as well as 10 of its states (Ministry of Environment, Forest and Climate Change, Government of India). The region extends 26°20′–35°40′ N and 74°50′–95°40′ E and covers the states Jammu and Kashmir, Himachal Pradesh, Uttarakhand, Sikkim, Arunachal Pradesh, Meghalaya, Nagaland, Manipur, and Mizoram, as well as the hill stations of Assam and West Bengal.

The region has been divided both horizontally and vertically into four and three divisions, respectively. In horizontal context, it has been categorized as Jammu and Kashmir Himalaya, Himachal Himalaya, Uttarakhand Himalaya, and Eastern Himalaya (Figure 1). Uttarakhand Himalaya is further divided into the Garhwal and Kumaon Himalayas. In terms of vertical divisions, it has been divided into Great Himalaya, Middle Himalaya, and Lesser Himalaya or Shivalik Ranges [5]. Geologically, the Himalayas are broadly divided into four areas across its length: (i) Foothill or Outer Himalaya; (ii) Lesser Himalaya; (iii) Higher Himalaya; and (iv) Tethyan or Trans-Himalaya [6]. The major rivers that originate from the great Himalayan Mountain Ranges are the Indus, Ganges, and their various tributaries. The region exposes diverse geology with different rock types representing the complete spectrum ranging in age from Archaean metamorphites/granitoids to the youngest Quaternary alluvium.

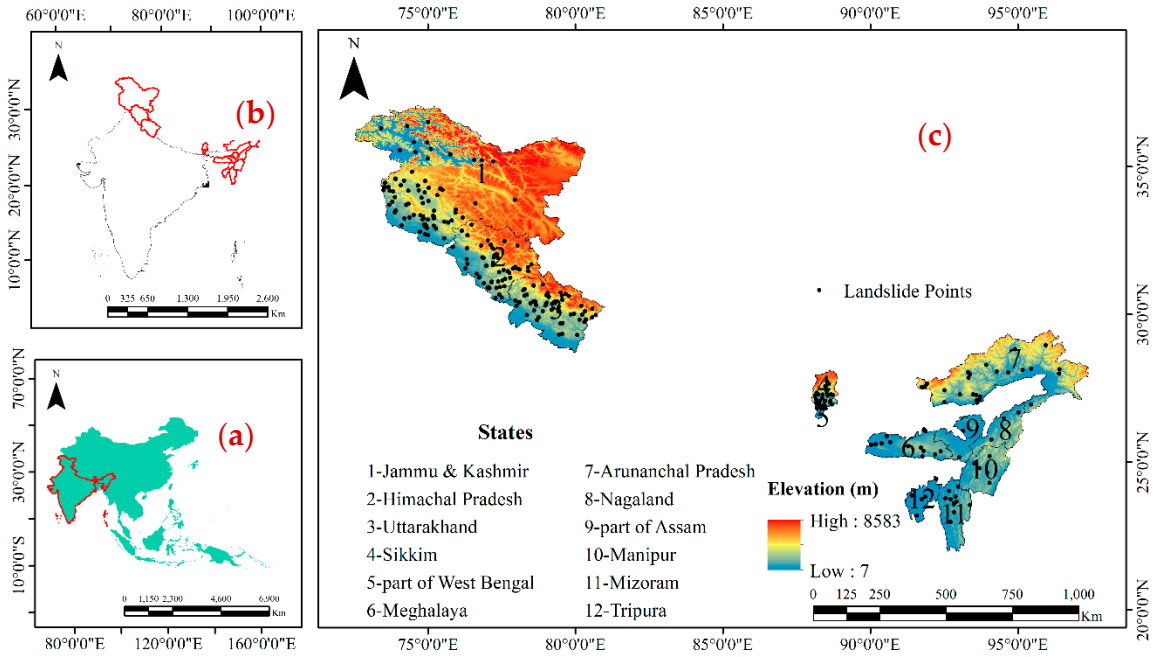


Figure 1. (**a**) Southeast Asia highlighting India; (**b**) geographical boundaries of India highlighting the Indian Himalayan region; and (**c**) location of rainfall-induced landslides in the Indian Himalayan region [1] (DEM Source: SRTM, resolution 30 m).

The region is tectonically and physiographically divided into three broad domains: the Peninsular India, the Extra-Peninsular India, and the Indo-Gangetic Plain [7]. The Indo-Gangetic plain is sandwiched between the shield area of the Peninsular India and the highly deformed suites of the Himalaya of Extra-Peninsular India, comprising essentially the younger meta-sediments. The tectonic trough (foreland basin) sandwiched between peninsular shield in the south and Himalayan Mountains in the north formed due to upliftment of the latter has been filled up by sediments derived from both sides, especially from the Himalayas. Structurally, the Himalayan Mountain chain occurring all along northern part of India can be divided into four contrasting longitudinal litho-cum-morphotectonic belts from south to north: (i) foothill belt; (ii) Main Himalayan belt; (iii) Indus-Shyok belt; and (iv) Karakoram belt.

The foothill Himalaya is a 10–50 km wide Miocene to Lower Pleistocene Molasse sequence represented by Siwalik, Murree, and Subathu Group of rocks. The belt is a domain of active tectonics, having participated in the terminal phase of the Himalayan Orogeny. This is followed to the north by the Lesser and Higher Himalayas, represented by geological sequences of Proterozoic age with a Phanerozoic cover of varying thicknesses in different parts. The foothill Himalaya is overlain by alluvium and separated from the Lesser Himalaya by the north-dipping fault commonly known as the Main Boundary Fault (MBF) or the Main Boundary Thrust (MBT) in Garhwal, Kumaon, Darjeeling, and Arunachal Pradesh Himalaya. The Main Frontal Thrust (MFT) limits the margins of the Siwalik Zone against the Ganga Plains [8].

The Lesser Himalaya is 60–80 km wide and is a discontinuous belt stretching between the MBT in the south and the Main Central Thrust (MCT) in the north [9]. It consists of autochthonous Late Proterozoic sediments, thrust over by three vast nappes that are built up successively of Palaeozoic sediments, Precambrian epimetamorphics, and mesograde metasediments [10]. The epi-metamorphic and meso-metamorphic nappes throughout their extent are characterized by Early Proterozoic (= 1900 Ma) and Early Palaeozoic granitic bodies of large dimension. The MCT separates the Lesser Himalaya from the Higher Himalaya to its north. The Higher Himalaya marks the region of the highest peaks of the Himalaya (Nunkun, Leopargial, Kedarnath, Badrinath, Nanda Devi, Api, Dhaulagiri, Mt. Everest, and Kanchanjunga), made up of 10–15 km thick Precambrian crystallines exhumed up and intruded by

granites, some of which are Tertiary in age. The Indus Shyok belt/the Tethys Himalaya extend to the south of the Trans-Himalayan Karakoram belt and comprise ophiolite melange (Indus ophiolite and associated formation) and plutonic rocks (Ladakh Granitoid Complex) of the Indus Shyok belt [11]. These predominantly fossiliferous sediments range in age from Late Proterozoic to Eocene. Sporadic occurrences of chromite have been reported from the ultrabasic rocks associated with Dras volcanics from Ophiolite-Melange zone. Karakoram belt, the northernmost zone, comprises Palaeozoic and Mesozoic sedimentary sequence in a metamorphic basement of an unknown age. This Trans-Himalayan belt lies to the north of the Indus suture Zonein Ladakh region and extends eastward into Tibet. No important mineral occurrence is known from this belt. Figure 2 represents the geological map of the Indian Himalayan region.

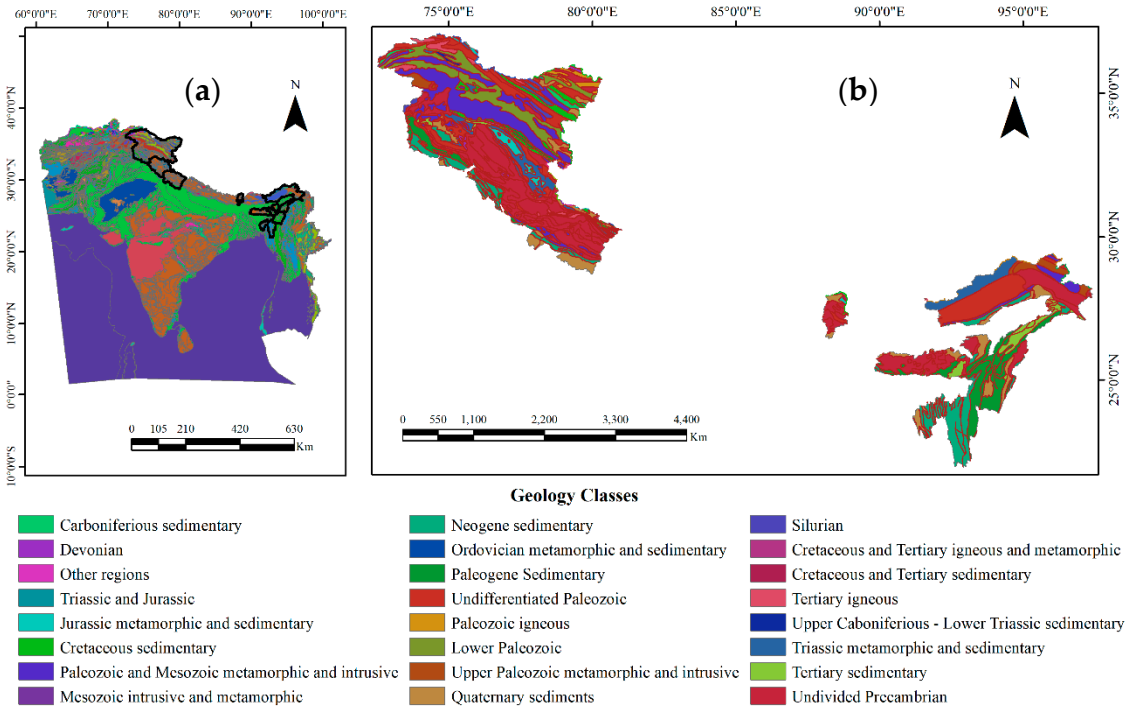


Figure 2. Geology of: (**a**) South Asia highlighting the Indian Himalayan region; and (**b**) the study region (Source: USGS World Geologic Map).

2.2. Data Collection

The review was undertaken by carrying out a bibliographic search on "Web of Science" Database (1990–October 2019) for a combination of keywords: "Landslide*", "Himalaya*", and "India". For the current analysis, only peer-reviewed journal articles written in English were considered, as peer-reviewed journals are considered as having the best quality articles and because the English language ensures that these contributions could be fully understandable for the whole international scientific community. The filtering of the articles was carried out by the number of citations (a) articles published before 2008 with 10 or more citations; (b) articles from 2008 to 2010 with 5 or more citations; and (c) all articles thereafter independent of the number of citations received (following the approach adopted from Reichenbach et al. [12]. After performing a screening of the relevant studies, the number of selected articles was narrowed down to 226 (Figure 3).

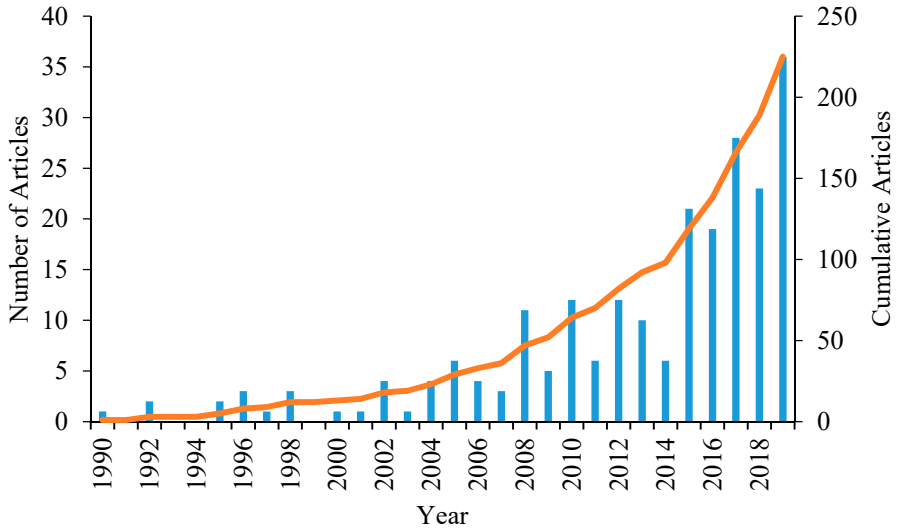


Figure 3. Analysis of the literature database during 1990–October 2019 from Web of Science. The left ordinate axis represents the number of articles per year and the right ordinate axis depicts the cumulative number of articles during the entire analysis period.

The articles were published across 75 different journals, with nearly 60% of the works published in 11 journals (Figure 4). The articles were initially majorly published in Engineering Geology and Geomorphology. The focus has shifted towards other journals such as Natural Hazards, Landslides, and Geomatics, Natural Hazards and Risk. Generally, after a major landslide event, the preliminary report is published in Current Science. However, extensive analysis of landslide study misses in such journals and is mostly focused on post-landslide studies.

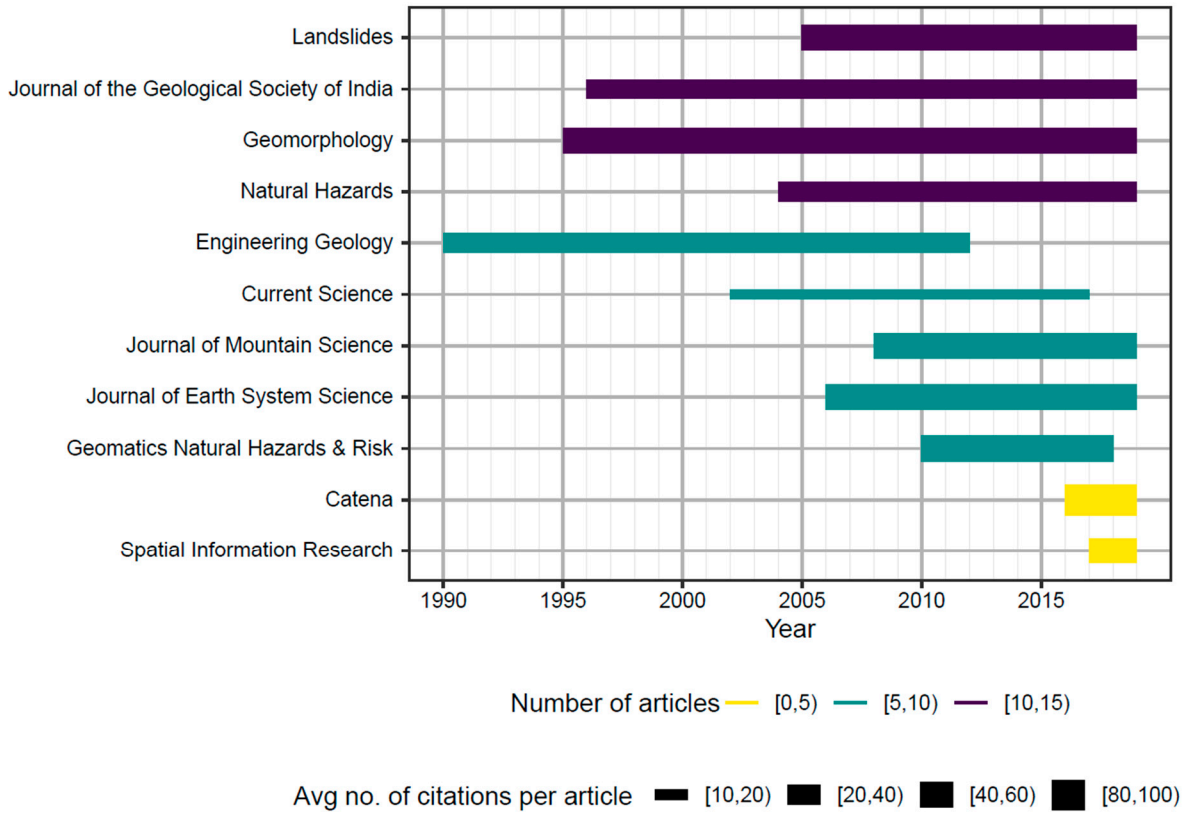


Figure 4. Distribution of the number of articles across the top 11 journals (out of 75). Horizontal bars depict the number of articles and its thickness denotes the average number of citations per article.

3. Results

The analysis reveals that the landslide studies in the region are quite biased towards Uttarakhand region, while there are few to no studies in the northeast Indian region (Figure 5).

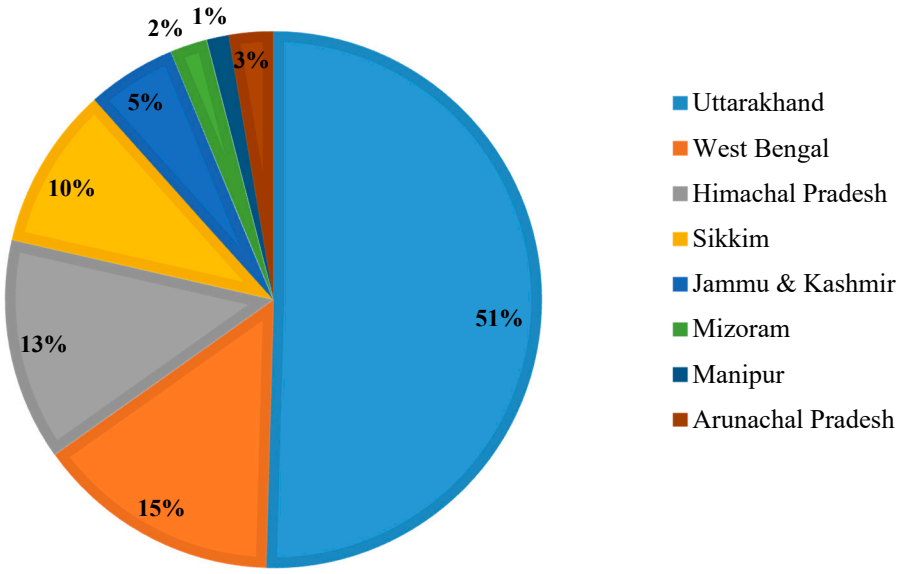


Figure 5. Distribution of landslide studies across the states in Indian Himalayan region.

This bias can be attributed to the high population density in the western sector of the Indian Himalaya, leading to a larger number of casualties and a generally higher level of landslide risk. Pham et al. [13] calculated the number of casualties due to landslide incidences during 2007–2015, which came out to be 5228, a staggering 82.9% of all disaster casualties in the region. To understand the type of studies being conducted, we divided them into seven broad categories: description of main events; identification; forecasting; monitoring and investigation; river damming; extreme events and climate change; and susceptibility mapping.

3.1. Main Landslide Events

The Himalayan region has been affected by several landslides, and a few of the most researched and damaging landslides are mentioned here. Most studies have been carried out in the Uttarakhand region. Some major landslide events in the Uttarakhand region occurred in the Okhimath area in Mandakini Valley, which was significantly damaged by the August 1998 landslide event, triggered by heavy rainfall. In total, 466 landslides were triggered, leading to 103 deaths and damaging 47 villages. Thereafter, cloudburst on July 2001 in a part of Mandakini valley led to more than 200 landslides causing 27 fatalities and affecting almost 4000 people [14]. The region also suffered from a heavy rainfall of more than 200 mm during 13–14 September 2012, which led to the death of 51 people and caused 473 landslides [15]. June 2013 witnessed another heavy rainfall of 350 mm within a period of five days in the Chamoli and Rudraprayag districts of Uttarakhand [16]. Other major landslides in the Uttarakhand region were the 2009 landslide in Pittorgarh region with a loss of 43 people.

Another major landslide affected region is Himachal Pradesh, of which the area of Katropi has often been studied [17,18]. The area has a history of recurring landslides since 1977 and recently it suffered from major landside in August 2017 claiming 46 lives [17]. The Pawari landslide zone has also been an active landslide zone located in the southeastern part of the state, which expanded 7% during 2005–2014 [19]. The Luggarbhati landslide in 1995 and Dharla landslide event in 2007 claimed 65 and 62 lives, respectively. The Jammu and Kashmir region has suffered from major cloudbursts such as the Leh Nalla cloudburst (2005, 2006), Leh cloudburst (2010), and Baggar cloudburst (2011) [20].

In the northeast part of the region, most reported landslides have been in the Darjeeling area of West Bengal and Sikkim. The Darjeeling region has a history of several landslides with the first major recorded occurrence in 1899, which led to the death of 72 people. Thereafter, landslides happened in 1950 (127 people died) and 1968 (667 people died) [21]. The region has also suffered significant landslides in the late 20th and 21st century (1991, 1993, 2004, 2005, 2006, 2009, and 2015) [22]. Similarly, the Sikkim region faced numerous landslides such as the Lanta Khola landslide [23], south Sikkim area [24]. The prominent unstable sections in the northeast region of the Indian Himalaya are National Highway (NH)-40 [25] and NH-44 [26] in Meghalaya. In Mizoram, NH-44A has suffered several landslides, in 2011, 2013 and 2017 [27].

3.2. Landslide Identification

The identification, visualization and standardized classification is an important step for pre- and post-hazard landslide analysis, which could assist in assessing rescue and relief operations [28]. Historically, landslide damage assessment was carried out using field visits and categorizing it according to the block diagrams depicted in [29]. However, understanding landslide scenario using such simplistic illustrations fails to consider the surrounding morphometry and its contextual relationship. With the advancement of remote sensing technologies and availability of high spatial and temporal resolution data, researchers have attempted to map landslides using image classification techniques. In the case of Indian Himalaya, the identification of landslides after an event has been mostly carried out by scientists at National Remote Sensing Centre (NRSC). Vinod Kumar et al. [30] studied Varunawat Parbhat landslide in Uttarakhand following the September 2003 landslide event using post-landslide Stereoscopic Earth Observation data on a 1:10,000 scale. The availability of high-quality images enables researchers to identify the loss of vegetation and exposure of fresh rock and soil after the landslide event [31]. The methodologies to assess landslide damages using pre- and post-landslide event images can be categorized as pixel based (PB) (medium resolution images) and object based (OB) (high-resolution images). Pixel based method involves the use of pre- and post-event images, and landslides are identified using spectral information. The identification of the landslides is mainly based on change detection and image fusion techniques. However, the use of pixel-based identification approach has some limitations. Firstly, the availability of images (pre- and post-event) may not be available from the same sensor, including variation in the atmospheric conditions and bandwidth of spectral values. Secondly, the representations of landslides are based on pixel values. On the other hand, object-based approaches use spectral, shape, spatial, and contextual properties and the identification is carried out using an object rather than a pixel, which makes more sense for natural events such as landslides as the damages are of irregular size and shape. It has the potential to add morphometric information derived from Digital Elevation Model (DEM).

The work conducted in the Indian Himalayan region has been primarily conducted using an object-based image classification technique. The use of OB image classification is dependent on the data type and methodology being used. The methodology involves detection, segmentation, and classification of images. The detection procedure can further be categorized into: (i) direct object change detection; and (ii) classified object change detection. The first detection technique involves the comparison of two images for changes in geometrical and spectral information acquired at different times. Such approach is usually used when comparison is made between images acquired from same satellite sensors and generation of full change matrix is not an absolute requirement. In the latter approach, first segmentation and classification are performed independently using time series images, and then the change matrix is derived using classified images. The second approach is mostly used in land use/land cover change analysis as well as when time series data are acquired by different satellite sensors [32]. The studies conducted in the region have mainly used the first approach as a single feature, i.e., the landslide is analyzed and the changes are estimated from images acquired from the same sensor.

The capability of detection using object-based analysis depends on the segmentation technique being used, which has the ability to excerpt objects that accurately describe the relevant properties [32]. The most popular segmentation technique for classification of images is the multi-resolution segmentation (MRS). This segmentation approach is a bottom-up region-merging technique, wherein small objects are merged into bigger ones in subsequent steps, and involves three parameters: scale, shape, and compactness. The scale parameter handles the size of the image object size, whereas the shape parameter determines the degree to which shape influences segmentation vs. spectral homogeneity. The compactness defines the weight of the compactness criteria. The higher the value is, more compact the objects will be [33]. Several methods exist to select an optimal scale parameter, of which plateau objective function [31,32] and optimal scale parameter selector [34] have mostly been used. However, obtaining a desired scale for all the landslide types in an area is difficult, thus over-segmentation is preferred to under-segmentation [35].

The use of the OB technique has been used for landslide detection immediately after a landslide event, as well as for creating a landslide inventory database from historical images. Martha et al. [35] prepared landslide inventory data for the 1998 Uttarakhand landslides in Okhimath region, Uttarakhand and compared with the field data acquired after the landslide event. The study identified 73 landslides using Resourcesat-1 LISS-IV multispectral data (5.8 m) and a 10-m Cartosat-1 derived DEM. This semi-automatic approach resulted in achieving an accuracy of 76% for recognition and 69% for classification in terms of number of landslides. Thereafter, Martha et al. [31] used historical panchromatic image dataset (1998–2006) for the same region to prepare a landslide inventory database. The images were acquired from Cartosat-1 (2.5 m) and IRS-1D (5.8 m) and depicted the use of texture in cases of missing spectral information. Martha et al. [36] analyzed various time series images (1997–2009) for Okhimath region, Uttarakhand using Cartosat-1 (2.5 m), Resourcesat-1 (5.8 m), and IRS-1D panchromatic (5.8 m) data. For the entire period, the accuracy of landslide detection varied from 60% to 89%, whereas it varied from 71% to 97% in terms of landslide extent.

The identification and classification of landslides is available for two severe landslide events occurring in Uttarakhand in September 2012 [15] and June 2013 [16,32]. Martha and Vinod Kumar [15] used very-high resolution (VHR) satellite data (Resourcesat-2 (5.8 m), Cartosat-2 (1 m), Kompsat-2 (1 m), GeoEye-1 (0.5 m), and Cartosat-1 DEM (10 m)) covering the September 12 Uttarakhand landslide event and identified 473 landslides. Martha et al. [16] compared pre-disaster images (Resourcecat-2 and GeoEye-1) with post-disaster images (Resourcesat-2 and Cartosat-2a (1 m)) after the June 2013 Uttarakhand event. The study identified 6031 landslides and classified those as new (57.74%), old (18.92%), and reactivated (23.34%). Martha et al. [32] identified new landslides by comparing the pre- and post-landslide Resourcesat-2 images with a detection accuracy of 81%. Mohan Vamsee et al. [34] improved the scale component of MRS technique and developed the optimal scale parameter selector (OSPS) tool, which was applied to Uttarakhand region using Resourcesat-2 images.

However, the accuracy of the above-mentioned techniques largely depends on the resolution of the satellite images that are costly to be acquired. Therefore, free and high-resolution Google Earth (GE) images have also been used for landslide mapping [19,37]. The ability of such images to provide a 3D view and its free availability can serve researchers to exploit imagery for landslide detection and mapping [38]. Kumar et al. [37] utilized GE images for landslide dimension mapping along Satluj Valley in Northwest Himalaya. Further, Kumar et al. [19] conducted a study on a relatively smaller area (Pawari landslide) in the same region and analyzed the changes using GE images of 2005, 2012, and 2014.

As landslide mapping is the first key step towards conducting any landslide study or to set up recovery attempts after a landslide event, focus should be on the use of high temporal resolution dataset. The focus has primarily been on the use of semi-automatic identification approach and gradually improving the algorithms used to detect and classify landslides [16,28]. The classification and segmentation techniques need to be improved using computational techniques such as Machine Learning (ML) and Artificial Neural Networks (ANN) [39]. In addition, the studies have largely been

concentrated in a single area and need to be applied to other Himalayan regions to understand their applicability and reliability.

3.3. Landslide Forecasting

Landslide forecasting is a key element for disaster risk reduction and is also the most challenging. Most of the landslides in the Indian Himalayan region are shallow in nature, of which rainfall is the primary triggering factor [3,4]. The analysis of precipitation for landslide occurrence can be performed by estimating minimum rainfall conditions, sub surface monitoring, or slope stability analysis. The minimum rainfall conditions, also known as thresholds, can be classified as empirical and physical approaches [40–42]. Physical models assess the relationship between rainfall conditions and hydrological conditions of the soil, which affects slope stability. The model analyzes the spatial variation of several factors such as geotechnical parameters, soil depth, volumetric water content, geology, and topography to determine the pore water pressure change and estimate the factor-of-safety. Such models need a large dataset of many parameters in spatial and temporal context, which is usually not available for the Indian Himalaya. Empirical methods analyze the rainfall conditions using statistical methods to determine threshold levels of precipitation. Such models are simpler to apply because they require only the spatial and temporal dataset of precipitation and landslide events [43,44]. The threshold values can be greatly affected by the data quantity and quality, the rain gauge density, and the methodology used. The thresholds are usually determined by drawing (usually with statistical techniques) a lower line to the precipitation conditions corresponding to landslide events in log-log, semi-log, or Cartesian coordinate system [3,41]. Segoni et al. [43] reviewed the type of thresholds, data collected, and other important information for thresholds determined in the global context for 2008–2017.

The studies on rainfall thresholds for the Indian Himalayan region have been very limited. The use of empirical approach has been explored for different Himalayan regions. Sengupta et al. [45] proposed the use of EMAP for Sikkim Himalayas, which is the ratio of cumulative event rainfall (E) to mean annual rainfall (MAP) (E_{MAP} = E⁄MAP), instead of rainfall event-duration (ED) or rainfall intensity-duration (ID) thresholds. Kanungo and Sharma [3] determined ID thresholds for Garhwal Himalayas using best fit of the lower boundary of ID plane. Dikshit and Satyam [4] developed the ID thresholds for Kalimpong region using frequentist statistical approach. The use of probabilistic [22,46] and semi-automated algorithmic approach [47] has only been attempted for Kalimpong region. Harilal et al. [48] developed both regional (Sikkim) and local (Gangtok, Sikkim) ID thresholds using statistical approach. The main concerns with these studies are the coarse spatial distribution of rain gauges, all studies being performed using a single rain gauge, and the availability of only daily rainfall data. In this regard, Gariano et al. [49] highlighted that a coarse temporal resolution dataset of precipitation could lead to an underestimation of rainfall thresholds, which culminates in a higher number of false alarms when using thresholds for operational early warning system. In addition, the thresholds determined are rainfall intensity-duration thresholds, which should be avoided, while the determination of event rainfall-duration (ED) thresholds should be encouraged. Mathew et al. [50] established ID thresholds along the Rishikesh–Mana pilgrimage route for Garhwal Himalaya using Tropical Rainfall Measuring Mission (TRMM)-based Multi-satellite Precipitation Analysis (TMPA) data. Despite using different methods to determine various aspects of rainfall characteristics, all studies were unified in the effect of antecedent rainfall in the region. Mathew et al. [50] also suggested that antecedent rainfall ranging from 15 to 30 days plays an influential role for destabilizing slopes in the Himalayan region, which leads the subsequent rainfall of short duration (24–72 h) to trigger landslides. The studies on rainfall thresholds in the Indian Himalayan region have been very minimal and more work should be conducted on the calculation and analysis of regional and local thresholds. Kumar et al. [51] highlighted that the mean rainfall threshold intensity of NW Himalayas (excluding Sikkim Himalayas) is roughly 290 mm/day. At present, defining rainfall thresholds at regional scale or at a state administrative level would be highly desirable but is hard to accomplish. The variation in the

thresholds determined also ascertains that local thresholds would perform be better than regional thresholds in setting up an operational landslide early warning system. This could be due to the heterogeneous rainfall pattern at local scales in various Himalayan pockets. Table 1 lists the various thresholds generated for different regions of the study area. Figure 6 illustrates the thresholds in ID plane.

Table 1. Threshold equations generated for various Himalayan regions using several methods (I is Rainfall Intensity (mm/h), while D is Duration (h)).

Threshold Equation	Region	Methodologies
$I = 1.82\,D^{-0.23}$	Chamoli, Uttarakhand	Empirical thresholds
$I = 58.7\,D^{-1.12}$	NH 58, Uttarakhand	Empirical thresholds
$I = 3.52\,D^{-0.41}$	Kalimpong, West Bengal	Empirical thresholds
$E = (4.2 \pm 1.3)\,D^{(0.56 \pm 0.05)}$	Kalimpong, West Bengal	Semi-automated algorithm approach
$I = 100\,D^{-0.92}$	Gangtok, Sikkim	Empirical thresholds
$I = 43.26\,D^{-0.78}$	Sikkim	Empirical thresholds

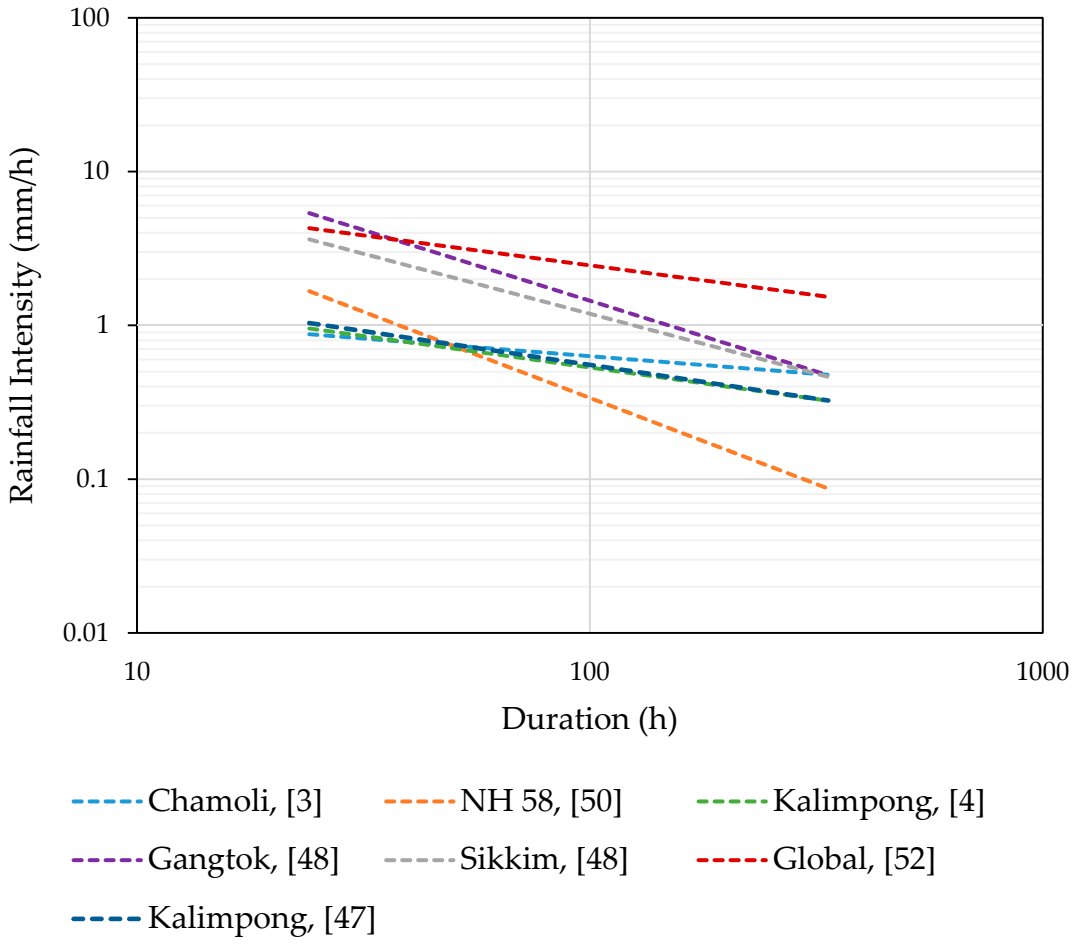


Figure 6. Comparison of various ID thresholds developed for various Himalayan regions with global thresholds defined by Caine [52].

3.4. Landslide Monitoring and Investigation

Landslide monitoring is a very significant aspect of landslide assessment, especially in the Indian Himalayan region where the slopes are generally creepy in nature, which could immediately fail during an abrupt rainfall event or seismic occurrence. Landslide monitoring is generally categorized into three types [53]:

1. Observing the changes of topography and cracks on the surface during the site investigation. This is a traditional monitoring technique used by geologists at regular time periods. The major limitation of using such a technique is that it does not provide the variations for a short time interval and it is difficult to accurately determine the time and location of future landslide incidences.
2. Remote sensing techniques such as satellite imagery analysis, GPS synthetic aperture radar (SAR) interferometry, and light detection and ranging (LiDAR). These methods can be helpful as they can measure slope displacement over a large area independent of the weather conditions.
3. In situ ground-based observation of slope movement using various instruments (extensometers, inclinometers, and tiltmeters) and installing rain gauges to accumulate local rainfall data. For example, Dikshit et al. [54] used tilt sensors for Darjeeling Himalayas, while Falae et al. [55] used Electrical resistivity tomography (ERT) for Garhwal Himalayas.

Of the various monitoring techniques, the use of remote sensing (RS) data along with traditional monitoring instruments would prove to be very helpful especially in understanding long-term deformation and the advancement in RS technology would gradually overcome the use of traditional equipment. In the case of Indian Himalaya, monitoring studies include remote sensing techniques and ground based observation. Yhokha et al. [56] used the Persistent Scatterer Interferometry (PSI) technique using ENVISAT satellite for Lesser Himalaya, Nainital. PSI is based on InSAR technique, which utilizes several SAR images and has proved to be successful in identifying the creeping zones in Nainital. Martha et al. [57] monitored the landslide dammed lake in Zanskar Himalayas for five months (January–May 2015) using multi-temporal high-resolution satellite images after the landslide event (December 2014) and depicted the variation in the dimension of the impounded lake during the monitoring period. Dikshit et al. [54] and Dikshit and Satyam, [22] used tilt sensors at shallow depths to analyze the variation in tilting angle of the instrument, which is related to the lateral displacement of the slope. The study also validated the empirical models thereby encouraging similar studies to be conducted in other Northeast Himalayan regions, which could help in setting up a preliminary early warning system.

The investigation of unstable slopes in the Indian Himalaya has been conducted using Ground Penetrating Radar (GPR) or 2D Electrical Resistivity Tomography (ERT). Mondal et al. [58] conducted ERT investigation for Naitwar Bazar landslide in Uttarakhand for six sites, which was active after the 2004 event. Kannaujiya et al. [59] compared slide dimensions observed from satellite (IRS LISS-IV) with geophysical investigation and determined the depth and slip surface geometry using 2D ERT and GPR) for Kaliganga river valley in Uttarakhand. Falae et al. [55] used ERT to understand the subsurface movement to determine the probable failure plane of the Pakhi landslide, Uttarakhand. Sharma et al. [60] conducted a geophysical study at Lanta Khola Landslide, Sikkim using very low-frequency (VLF) electromagnetic survey to understand the subsurface structure. The results reveal the presence of a water-saturated zone at the subsurface level of the slide.

The literature review reveals that only a handful of studies have been carried out for landslide monitoring and investigation; considering the increased quantity of the available remote sensing data sources, the focus needs to shift towards the use of high spatial and temporal resolution data, which would lead to near real-time monitoring results. The framework using sub surface investigation is ideal at a local scale and would fulfill the demands of a community but would be difficult to manage in long term due to the high installation and operational costs. However, in sections where the landslide problem is immense, which is slide specific, such as Paglajhora landslide (Darjeeling), Singtam landslide (Sikkim), or Kotropi landslide (Himachal Pradesh), monitoring using instruments would still be required as the region is highly vulnerable and satellite images at currently available resolution may not be able to accurately identify failure planar sections.

3.5. Lake Damming Landslides

Landslide lakes or dams are temporary lakes in the river valleys formed after a landslide blocks the river course. Landslide dammed lakes and their outburst floods (LLOFs) are common in the Indian

Himalaya. The breaching of such temporary lakes with a huge amount of accumulated water and sediments can create devastating floods in the downstream areas [61]. In the Himalayas, landslide dams commonly form in high mountains because of different mass movement types such as rock and debris avalanches, rock and soil slides, mud–debris–earth flows. The oldest recorded landslide lake was the Gohna Tal, which was formed in September 1893 as a result of heavy precipitation blocking the Birahi Ganga River situated in the Kumaon Himalaya. Gradually, the lake expanded to an approximate 4000 m length, 340 m width, and 300 m depth. On 26 August 1894, the dam collapsed, and a devastating flood hit downstream of Birahi Ganga [62].

The techniques used for assessment of damming processes are either based on geomorphic characteristics, hydraulic properties, and velocity measurement of dam material. The assessments of the damming process are dependent on the analysis of three significant factors: (a) pre-dam formation mechanism; (b) dimensional characteristics of the dam; and (c) stability analysis of dam [63]. The pre-dam formation mechanism largely depends on the slope stability analysis and landslide triggering factors, whereas dimensional characteristics are dependent on geometry of the area and landslide volume. In terms of slope stability, the techniques can be categorized as discontinuum modeling and continuum modeling [63]. Discontinuum modeling is majorly used for rock slope stability analysis, whereas continuum modeling is used for debris flows as well as rock slopes dependent on the material and geometry of the slope [19,64].

Most of the studies have focused on geomorphic analysis [15,16,57,63] using different satellite images to understand the spatiotemporal landslide changes. Gupta and Sah [65] catalogued the LLOFs developed in the Trans-Himalayan region between 2000 and 2005 and studied its impact on the stability in the region. Martha et al. [57] conducted an extensive investigation of Phutkal River landslide dammed lake for landslide occurrence on 31 December 2014 using multi-temporal Cartosat-2 images of 1 m spatial resolution and calculated slide volume using pre- and post-event datasets. Kumar et al. [63] used GE images of 1.5 m resolution and estimated landslide volume along with slope stability analysis for Urni landslide in Himachal Pradesh.

3.6. Extreme Events and Climate Change

The effect of global warming and the corresponding changes to climate and geohazards is expected to affect landslide events [66]. However, forecasting and understanding the impact of climate change on landslide activity still poses a challenge. Gariano and Guzzetti [66] in their review article on climate change studies related to landslides emphasized the need for more studies as large parts of the world suffer from a few to no studies. Although it is predicted that the Indian Himalaya region will be affected by climate change at the time of compiling this article, we could not find any article in which the effect of climate change has been considered in the Indian Himalaya for any type of rainfall-induced landslide study. However, some works exist that describe some of the most recent extreme rainfall events: Jammu and Kashmir Himalaya for the 2014 rainfall event [51], Uttarakhand Himalaya for the 2009 event [67], and Leh region for the 2010 event [68]. One of the main reasons for the lack of inclusion of climatic variation in landslide studies for the Indian Himalayan region is the unavailability of data both in spatial and temporal context, which has been highlighted by several researchers. Kumar et al. [51] estimated that the pattern of rainfall intensity along the Himalayas varies from west to east, thereby precipitation patterns being affected by western disturbances and summer monsoon. Such lack of studies in a region such as the Himalayas, which is prone to severe landslides and other geohazards, is a matter of concern. Hereon, studies should include climate aspects using remote sensing products, which could reconstruct events and provide a better understanding of climate impact.

3.7. Landslide Susceptibility

Landslide hazard is defined as the probability of landslide occurrence of a specific type (“when” or “how frequent”) and magnitude (“how large”) [12]. The spatial probability of landslide hazard is assessed by carrying out landslide susceptibility mapping or popularly known as landslide hazard

zonation mapping. Landslide susceptibility assessment determines potential of landslide event considering several predisposing factors and investigating their spatial distribution [12]. The spatial occurrence of landslides is controlled by several and sometimes inter-playing factors such as geological setting, rainfall, morphology, soil, and vegetation conditions, thus landslide susceptibility assessment is not a straightforward task, and many methodologies have been developed to assist the susceptibility analysis. The classification techniques used for landslide susceptibility models has also changed over time. We categorized them into four groups: Qualitative, Semi-quantitative, Quantitative, and Deterministic (Figure 7). In this section, we explore the models, the parameters used, and the trend in susceptibility analysis.

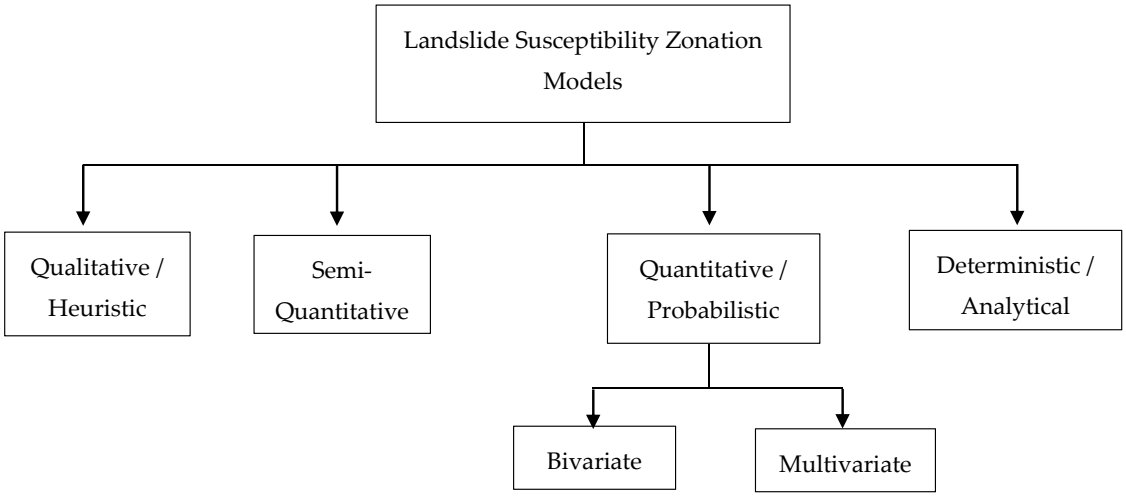


Figure 7. Types of modeling techniques used in landslide susceptibility studies.

Initially, studies were mostly based on qualitative methods wherein values are ascertained for landslide conditioning factors based on the knowledge about the study area and are quite subjective [69,70]. Semi-quantitative methods are logical tools and emphasize the significant factors by assigning weights or values. These include techniques such as Analytic Hierarchy Process (AHP), danger pixel approach, and a weighted linear combination [71].

The deterministic analysis involves an analysis of physical and mechanical soil properties and determines the susceptibility in the form of a factor of safety (FS) [72]. FS is the ratio between factors affecting landslide to factors preventing landslide incidences. This involves the utilization of several factors such as infiltration, soil cohesion, groundwater table, pore water pressure, geotechnical soil parameters, etc. Several deterministic models are available to determine slope stability such as SHALSTAB (Shallow Landsliding Stability model), SINMAP (Stability Index Mapping), SHETRAN, TRIGRS (Transient Rainfall Infiltration and Grid-based Regional Slope Stability) model, etc. In the case of the Indian Himalayas, Sarkar et al. [72] used SHALSTAB model to determine the critical steady-state rainfall for slope stability in Darjeeling Himalayas. A landslide inventory map was developed using various satellite images (Resourcesat -1, Cartosat -1, and Google Earth) and field surveys, soil parameters (depth, saturated soil density, engineering properties, and soil transmissivity), and slope parameters (angle, contour length, and upslope contributing area). Mathew et al. [73] determined FS for terrain stability for Garigaon watershed area of Uttarakhand by coupling an infinite slope stability model with a steady-state hydrological model (LIDA) using the spatial analysis in GIS environment. The landslide locations were mapped using LISS IV and Cartosat-1 PAN data and the input parameters used were soil texture, index properties, porosity, LULC, and terrain properties. The results indicate an increase in the unstable areas of more than 45% for rainfall intensity variation from 50 to 100 mm/day.

Quantitative models depend on the landslide density under each influencing factor and can be further classified as a bivariate and multivariate [74]. Both bivariate and multivariate statistical models compute weights; however, multivariate techniques depend on the collective effect of parameters [75,76]. Bivariate statistical methods include weights of evidence, frequency ratio, information value, and the combination of frequency ratio and fuzzy methods [77–79]. The most popular multivariate model used is logistic regression [80]. The analysis shows that most studies have used quantitative techniques, of which logistic regression has been most used [81–83]. Thereafter, frequency ratio and information value were used 16% and 9% of the studies, respectively [74,84–88]. In terms of semi-quantitative models, AHP has been the most used method (Figure 8).

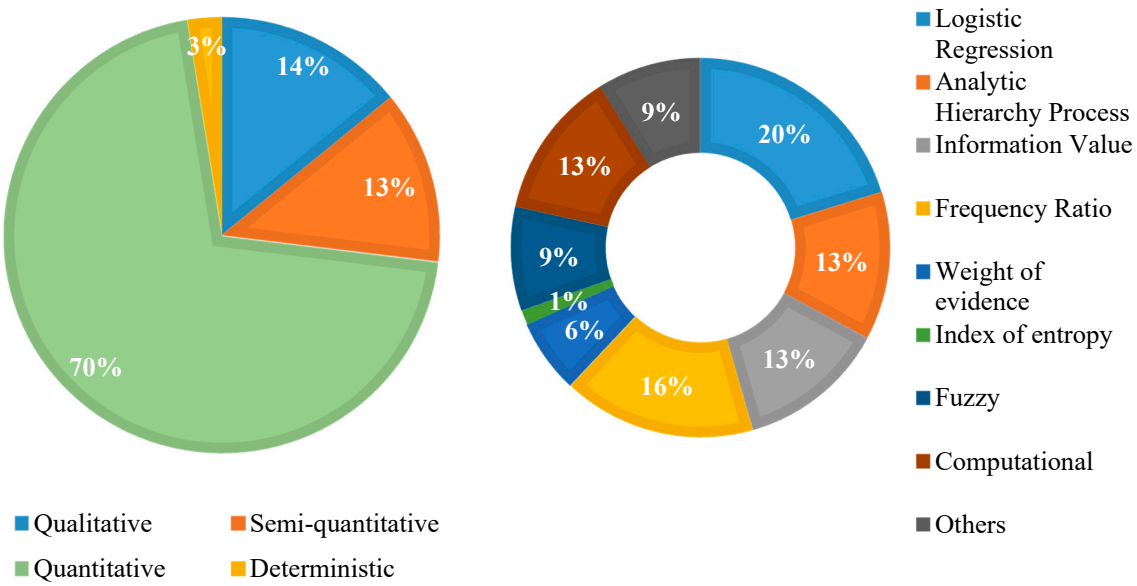


Figure 8. Model types and the techniques used in hazard and susceptibility studies.

After understanding the various models being used to determine susceptibility, the analysis of the parameters being used was investigated. The use of parameters for susceptibility analysis would affect the results; however, they usually depend on the local factors along with data availability. In addition, the factors depend on the type of movement, scale of study and the methodology used [89]. Based on the analysis of the articles, the factors influencing landslide susceptibility can be divided into four groups: (i) geological (lithology, geological structures); (ii) geomorphological (drainage, relative relief, slope, and slope aspect); (iii) environmental (land cover); and (iv) anthropogenic (roads). In the 81 studies, 36 factors were used, where some factors such as "gradient" and "slope" were clubbed together, and a word map of the factors is illustrated in Figure 9. The most used factors were slope angle (93.8%), land use/land cover (LULC) (92.5%), aspect (80.2%), and lithology (62.9%). Apart from these factors, another set of parameters, namely drainage density, curvature, topographic wetness index (TWI), stream power index (SPI), relative relief, lineament distance, and lineament density, was also utilized. Seismic factors, fault buffer, and road buffer have been less frequently used. The trend over the years in the number of factors to be used has increased with an average of six factors being used until 2010 and an average of eight factors thereafter. This trend can be ascertained to the increase in the availability of data and remote sensing products.

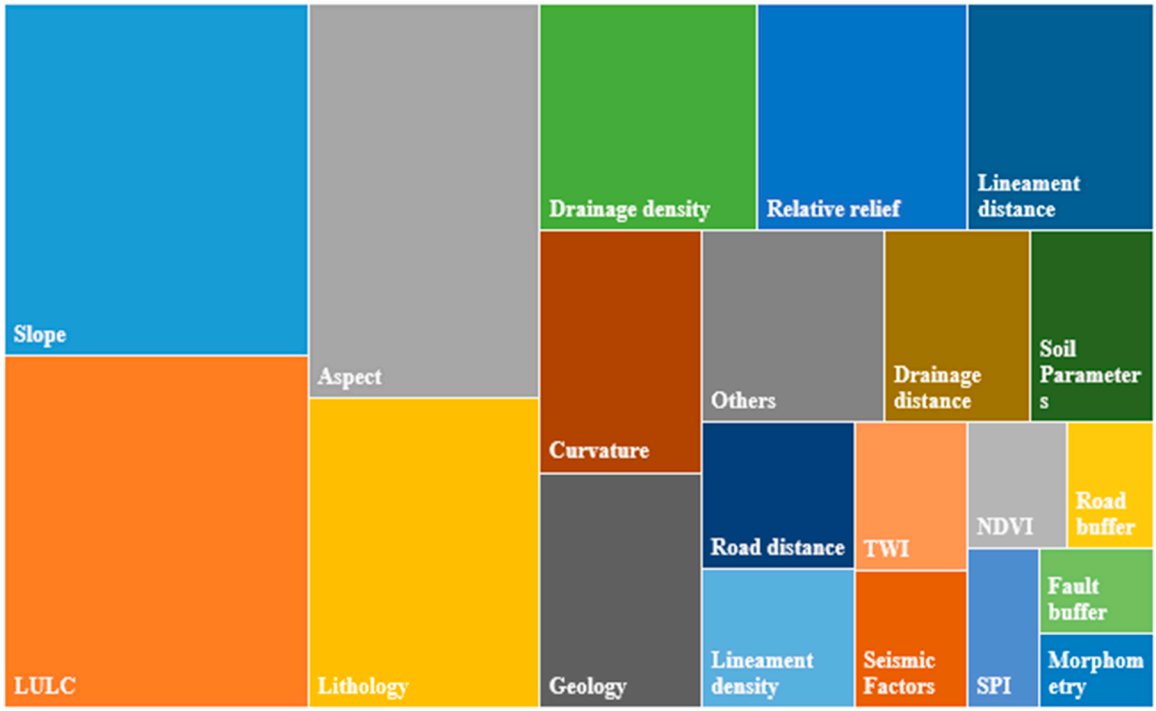


Figure 9. Tree map of the factors used for determining landslide hazard and susceptibility.

In terms of the spatial resolution of the dataset used to determine susceptibility, the most commonly used Digital Elevation Model (DEM) is the one released by USGS, with a spatial resolution of 30 m (46%). Only 28% of the studies used a spatial resolution higher than 10 m, which was obtained from Cartosat-1 images, whereas 26% of studies used the dataset with a spatial resolution of 12.5 m from ALOS-PALSAR DEM. The scale of geological maps has generally been 1:250,000 (62%) while the remaining 38% studies were conducted using 1:50,000 scale. Anbalgan [69] formulated the guidelines for landslide hazard assessment based on Landslide Hazard Evaluation factor (LHEF) using factors such as structure, relative relief, lithology, slope morphometry, land use, and groundwater conditions. The LHEF rating system was based on determining the factors and assigning values based on the understanding/knowledge about the study region. The national standard body of the country, Bureau of Indian Standards (BIS), prepared its guidelines (BIS, 1998) based on Anbalgan's study using the same factors excluding groundwater conditions [90]. BIS recommends an indirect method for medium-scale (1:25,000–1:50,000) landslide susceptibility mapping. Such technique is a generalized method that could be applied over large regions regardless of the relationship between causative factors and landslide types, often leading to moderate to poor predictions [91].

Following this, the nodal agency, Geological Survey of India (GSI), developed its own guidelines (GSI, 2005) considering 10 factors, which is a modified version of BIS (1998). Singh et al. [90] compared the zonation maps prepared using BIS and GSI guidelines for Arunachal Pradesh and found that GSI guidelines were better than BIS guidelines. Das et al. [92] studied homogenous susceptible units (HSU)-based landslide hazard utilizing spatiotemporal data and landslide size, which could better represent homogeneous susceptible regions. The study was conducted using high-resolution data for Northern Uttarakhand region. Martha et al. [36] applied the weight of evidence model to determine the susceptibility using a semi-automatic derived landslide inventories. Sarkar et al. [93] was the first to introduce landslide intensity as a parameter for hazard determination, which was applied for Garhwal Himalayas. The intensity was determined by analyzing the volume and velocity, which depends on the landslide area, debris thickness and the types of failures.

However, lately, the rise has been in the use of computational techniques such as Machine Learning (ML), Support Vector machine (SVM), and Artificial Neural Network (ANN), which have

proved to outperform traditional approaches [13,94–96]. However, the studies have largely focused on landsides in Uttarakhand Himalayas. Ramakrishnan et al. [97] used backpropagation neural network and showed prediction capability of 80%. Pham et al. [98] used various ML ensemble models using Multilayer Perceptron (MLP) Neural Networks and six different ensemble techniques, thereby depicting better capability using ensemble framework. Pham et al. [99] analyzed susceptibility for an area of 323 km^2 in Uttarakhand using various ML models (SMO-SVM, VFI, and LR), FT, MLP-neural networks, and NB. In the first case, SMO-SVM performed best, whereas, in the latter case, MLP neural networks and FT models provided similar accuracy compared to NB. Pham et al. [13] used four hybrid machine learning models for Uttarakhand Himalayas. Pham et al. [13,95,100] extensively analyzed the Pittorgarh region (242 km^2) in Uttarakhand using various hybrid ML models (Table 2).

Table 2. Various computational models used for landslide susceptibility in Uttarakhand Himalayas.

Models	No of landslides	No of Factors	Test Site Area/Pixel Size	AUC/Accuracy	Articles
ANN	154	7	600 km^2/5 m × 5 m	0.84	Chauhan et al. [101]
Back propagation neural network	63	6	~116 km^2/30 m × 30 m	0.8	Ramakrishna et al. [97]
SVM, Proximal SVM, L2-SVM-Modified Finite Newton	2009 pixels	8	1625 km^2/30 m × 30 m	0.807	Kumar et al. [102]
Multilayer Perceptron (MLP) Neural Networks	930	15	1325.47 km^2/20 m × 20 m	0.886	Pham et al. [94]
sequential minimal optimization (SMO) SVM, vote feature interval (VFI), Logistic Regression (LR)	430	11	323.82 km^2/20 m × 20 m	0.891	Pham et al. [98]
Functional Trees (FT), Multilayer Perceptron (MLP) Neural Networks, Naïve Bayes (NB)	430	11	0.32 km^2/20 m × 20 m	0.850	Pham et al. [99]
Rotation Forest based Radial Basis Function (RFRBF) neural network	930	15	0.13 km^2/20 m × 20 m	0.891	Pham et al. [103]
Aggregating One-Dependence Estimators, SVM, ANN-RBF, LR, NB	1295	16	561 km^2/30 m × 30 m	0.968	Pham et al. [104]
Ensemble decision tree	103	10	242 km^2/20 m × 20 m	0.883	Pham et al. [13]
Rotation Forest Ensemble Model	103	10	242 km^2/30 m × 30 m	0.741	Pham et al. [95]
Hybrid Reduced Error Pruning Trees	103	10	242 km^2/30 m × 30 m	0.989	Pham et al. [100]
Hybrid model of Multi Boost ensemble and SVM	391	16	270 km2/10 m × 10 m	0.972	Pham et al. [105]
Neuro-fuzzy inference system	391	16	270 km^2/30 m × 30 m	0.95	Jaafari et al. [106]
Fuzzy Expert System, Extreme Learning Machine	49	8	43 km^2/15 m × 15 m	0.844	Peethambaran et al. [96]

Kanungo et al. [107] compared susceptibility map for Darjeeling Himalayas using traditional, ANN, fuzzy, and combined neural fuzzy weighing schemes and found the combined neural fuzzy model to provide more accurate results. Following this, two different models (combined neural certainty and fuzzy) were compared with the neural fuzzy model for the same region and the latter was found to be better [108]. Chawla et al. [109] used Genetic Programming (GP) and Particle Swarm Optimization (PSO) SVM model for the Darjeeling Himalayan region and found GP method to perform better. Meena et al. [110] conducted susceptibility analysis using a hybrid spatial multicriteria

evaluation model for Kullu, Himachal Pradesh and found it to produce better results when compared to FR and AHP.

4. Discussion

The results of the review provide researchers a comprehensive insight to understand the studies being conducted in various avenues of the Indian Himalaya, thus providing a general picture of the state-of-the-art in this region and a reference point for those (researchers and practitioners) starting to work in this area for the first time. Moreover, the analysis of the results can be useful in identifying some existing research problem and the most important research questions to be addressed in the near future.

The first gap to be urgently filled is the lack of studies in regions such as the Northeast and Jammu and Kashmir Himalayas. Since research programs have focused mainly on Uttarakhand region, which is understandable given the high population density and risk exposure, more efforts need to be made to analyze other regions as well and to help policy makers in formulating a general guideline towards mitigation. Another issue is the urgent need for high quality data, as many of the reviewed studies suffer from the coarse dataset being employed. As an instance, rainfall thresholds have been conducted using a very sparse rain gauge network that measures rainfall only at daily time steps; and most of the susceptibility studies make use of a very coarse geological map at the 1:250,000 scale. It is well established that the poor quality of input data has a negative impact on the quality of the results of any scientific model; therefore, priority should be given to the acquisition of high-resolution basic data for further research.

In the Indian Himalaya, many studies have been conducted that explore different disciplines of landslide research in various regions, and a precise framework is not evident. One of the next steps of research in the region could be the combination of existing studies in a more comprehensive and coherent framework aimed at assessing and reducing landslide risk. For instance, susceptibility maps could be used to define the most hazardous areas, rainfall thresholds to provide temporal forecasts especially for rapid movements, regional scale remote sensing systems to get a general overview of slow-moving landslides, and ground based monitoring systems could be employed for management of the hot-spots at highest risk level. Similar programs should also consider landslide hazard evolution as influenced by climate change and human modifications to the territorial setting.

Another significant gap highlighted by the review is that, even though landslides are a well-known source of risk in the area, a quantitative risk assessment has never been carried out, thus all past and present risk reduction strategies are implemented without a comprehensive framework that could optimize the efforts made. At present, landslide hazard studies are also quite limited: examples exist that encompass temporal probability of occurrence (by means of rainfall thresholds) or spatial probability of occurrence (by landslide susceptibility maps). A spatiotemporal hazard assessment would be desirable and, to obtain it the future, two options could be explored: the application of distributed physically-based models (which, in turn, require the acquisition of many input parameters with high spatial density—see, e.g., Tofani et al. [111]) or the adoption of simpler approaches based on the dynamic combination of rainfall thresholds and susceptibility maps [112].

Several techniques have been used to understand various aspects; however, the use of computational techniques has been limited to susceptibility and hazard analysis. It is encouraging to see the use of advanced machine learning techniques (ANN, SVM, and Random Forest) for landslide mitigation. However, the use of deep learning and artificial intelligence, which shows great promise in geohazards and climatic studies, is yet to be performed for the Indian Himalayan region. The use of advanced computational techniques (deep learning) especially for landslide identification is one aspect which researchers should look to exploit.

5. Conclusions

This review paper is an attempt to understand the studies being conducted in the Indian Himalayan region, which contributes almost 15% of the global rainfall-induced landslides. The article focuses only on landslide activity which was triggered due to rainfall as it reflects a significant portion to landslides in the Indian Himalaya and it is the only typology that can be managed by early warnings and forecasting models. The review reveals that there are several avenues where significant progress is required, especially in the aspect of climate change, use of high spatial and temporal resolution data, and new methodologies such as physical methods and computational approaches. The study highlights key topics in which the research has focused: susceptibility mapping, identification, and slope stability analysis. The conclusions from each sub-section are as follows:

1. Landslide assessment (including identification, threshold estimation, and monitoring of landslides). For landslide identification and mapping, focus needs to shift in three key directions:
    a. Use of automated approaches involving the use of computational techniques.
    b. Use of higher temporal resolution datasets and assessment of their reliability.
    c. Application of the current techniques towards other significant landslide prone Himalayan regions. In terms of rainfall threshold studies, the number of articles were less than 10 with most of the work focused on the use of statistical models to define a single threshold, and majorly the Eastern Himalayas have been covered. The thresholds developed show large differences when calculated for regional and local scale, therefore it is suggested to develop thresholds at local scale to improve the understanding of the region and help in setting up an operational landslide early warning system. In general, the thresholds are very low if compared with other literature thresholds, thus confirming the basic assumption that the India Himalaya is very susceptible to landsliding. Additional research needs to be conducted on the use of physical models, including campaigns aimed at gathering input data for more complex models; moreover, efforts need to be made on the combination of empirical and physical models for a better understanding of landslide initiation.
2. Landslide monitoring has been performed using both ground instrumentation and satellite data, however a general multi-scale approach that extensively covers the whole region is missing.
3. The focus needs to shift on include climatic factors for landslide assessment as climate change is unequivocal. The use of climate models needs to be conducted with caution, especially when downscale projections are considered. As the climatic conditions are quite varied from west to east, focus should also be on the use of appropriate down-scaling models.
4. The studies on landslide susceptibility was found to have a regional bias and more research needs to be conducted in the Jammu and Kashmir Himalayas and the northeastern belt. Emphasis has primarily been on specific states and regions such as Uttarakhand, Darjeeling, and some areas of Himachal Pradesh. In modeling aspects, the use of computational approaches needs to be emphasized as it has proved to be better than traditional methods. The analysis should start focusing on the use of hybrid models and big data analytics for regional to site specific analysis, thereby understanding the heterogeneity and uncertainty of the region.
5. There is a serious lack of ground-based rainfall data in large parts of the Himalayan region which has been highlighted by several works. A solution to this flaw is to start using remote sensing data (e.g., satellite radar rainfall estimates) to compare and find the best dataset to be used for individual sections in the Himalayan region.

Author Contributions: Conceptualization, A.D. and R.S.; formal analysis, A.D. and S.S.; data curation, R.S.; writing—original draft preparation, A.D. and S.S.; writing—review and editing, B.P. and A.M.A.; and supervision, B.P. and R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) in the University of Technology Sydney (UTS) under Grants 321740.2232335, 321740.2232357,

321740.2232424, and 321740.2232452. This research was also funded by International Science Council under Grant ISCROAP/IRDR/SG/2019/009. This research was also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia. The research was also funded by Università degli Studi di Firenze under the Grant 58517_INTERNAZIONALIZZAZIONE.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Froude, M.J.; Petley, D. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
2. Dubey, C.S.; Chaudhry, M.; Sharma, B.K.; Pandey, A.C.; Singh, B. Visualization of 3-D digital elevation model for landslide assessment and prediction in mountainous terrain: A case study of Chandmari landslide, Sikkim, eastern Himalayas. *Geosci. J.* **2005**, *9*, 363–373. [CrossRef]
3. Kanungo, D.P.; Sharma, S. Rainfall thresholds for prediction of shallow landslides around Chamoli-Joshimath region, Garhwal Himalayas, India. *Landslides* **2014**, *11*, 629–638. [CrossRef]
4. Dikshit, A.; Satyam, D.N. Estimation of rainfall thresholds for landslide occurrences in Kalimpong, India. *Innov. Infrastruct. Solut.* **2018**, 3. [CrossRef]
5. Sati, V.P. *Towards Sustainable Livelihoods and Ecosystems in Mountain Regions*; Springer: Berlin, Germany, 2014.
6. Chakrabarti, B.K. Chapter 1—Lithotectonic Subdivisions of the Himalaya. In *Geology of the Himalayan Belt*; Chakrabarti, B.K., Ed.; Elsevier: Amsterdam, The Netherlands, 2016; pp. 1–9. [CrossRef]
7. Roy, A.B.; Purohit, R. Chapter 2—Indian Subcontinent: Geomorphic and Geophysical Traits. In *Indian Shield*; Roy, A.B., Purohit, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; pp. 13–30. [CrossRef]
8. Mukhopadhyay, D.K.; Mishra, P. The Main Frontal Thrust (MFT), northwestern Himalayas: Thrust trajectory and hanging wall fold geometry from balanced cross sections. *J. Geol. Soc. India* **2004**, *64*, 739–746.
9. Pradhan, B.; Singh, R.; Buchroithner, M. Estimation of stress and its use in evaluation of landslide prone regions using remote sensing data. *Adv. Space Res.* **2006**, *37*, 698–709. [CrossRef]
10. DeCelles, P.G.; Carrapa, B.; Gehrels, G.E.; Chakraborty, T.; Ghosh, P. Along-strike continuity of structure, stratigraphy, and kinematic history in the Himalayan thrust belt: The view from Northeastern India. *Tectonics* **2016**, *35*, 2995–3027. [CrossRef]
11. Thakur, V.C. Plate tectonic interpretation of the western Himalaya. *Tectonophysics* **1987**, *134*, 91–102. [CrossRef]
12. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
13. Pham, B.T.; Shirzadi, A.; Shahabi, H.; Omidvar, E.; Singh, S.K.; Sahana, M.; Asl, D.T.; Bin Ahmad, B.; Quoc, N.K.; Lee, S. Landslide Susceptibility Assessment by Novel Hybrid Machine Learning Algorithms. *Sustainability* **2019**, 4386. [CrossRef]
14. Naithani, A.K.; Kumar, D.; Prasad, C. The catastrophic landslide of 16 July 2001 in Phata Byung area, Rudraprayag district, Garhwal Himalaya, India. *Curr. Sci.* **2002**, *82*, 921–923.
15. Martha, T.R.; Vinod Kumar, K. September, 2012 landslide events in Okhimath, India—An assessment of landslide consequences using very high resolution satellite data. *Landslides* **2013**, *10*, 469–479. [CrossRef]
16. Martha, T.R.; Roy, P.; Govindharaj, K.B.; Kumar, K.V.; Diwakar, P.G.; Dadhwal, V.K. Landslides triggered by the June 2013 extreme rainfall event in parts of Uttarakhand state, India. *Landslides* **2015**, *12*, 135–146. [CrossRef]
17. Roy, P.; Martha, T.R.; Jain, N.; Kumar, K.V. Reactivation of minor scars to major landslides–a satellite-based analysis of Kotropi landslide (13 August 2017) in Himachal Pradesh, India. *Curr. Sci.* **2018**, *115*, 395. [CrossRef]
18. Pradhan, S.P.; Panda, S.D.; Roul, A.R.; Thakur, M. Insights into the recent Kotropi landslide of August 2017, India: A geological investigation and slope stability analysis. *Landslides* **2019**, *16*, 1529–1537. [CrossRef]
19. Kumar, V.; Gupta, V.; Jamir, I. Hazard evaluation of progressive Pawari landslide zone, Satluj valley, Himachal Pradesh, India. *Nat. Hazards* **2018**, *93*, 1029–1047. [CrossRef]
20. Banerjee, A.; Dimri, A.P. Comparative analysis of two rainfall retrieval algorithms during extreme rainfall event: A case study on cloudburst, 2010 over Ladakh (Leh), Jammu and Kashmir. *Nat. Hazards* **2019**, *97*, 1357–1374. [CrossRef]

21. Mondal, S.; Mandal, S. Landslide susceptibility mapping of Darjeeling Himalaya, India using index of entropy (IOE) model. *Appl. Geomat.* **2019**, *11*, 129–146. [CrossRef]
22. Dikshit, A.; Satyam, N. Probabilistic rainfall thresholds in Chibo, India: Estimation and validation using monitoring system. *J. Mt. Sci.* **2019**, *16*, 870–883. [CrossRef]
23. Anbarasu, K.; Sengupta, A.; Gupta, S.; Sharma, S.P. Mechanism of activation of the Lanta Khola landslide in Sikkim Himalayas. *Landslides* **2010**, *7*, 135–147. [CrossRef]
24. Bera, A.; Mukhopadhyay, B.P.; Das, D. Landslide hazard zonation mapping using multi-criteria analysis with the help of GIS techniques: A case study from Eastern Himalayas, Namchi, South Sikkim. *Nat. Hazards* **2019**, *96*, 935–959. [CrossRef]
25. Umrao, R.K.; Singh, R.; Sharma, L.; Singh, T. Soil slope instability along a strategic road corridor in Meghalaya, north-eastern India. *Arab. J. Geosci.* **2017**, *10*, 260. [CrossRef]
26. Sarkar, K.; Buragohain, B.; Singh, T. Rock slope stability analysis along NH-44 in Sonapur area, Jaintia hills district, Meghalaya. *J. Geol. Soc. India* **2016**, *87*, 317–322. [CrossRef]
27. Sardana, S.; Verma, A.; Singh, A. Comparative analysis of rockmass characterization techniques for the stability prediction of road cut slopes along NH-44A, Mizoram, India. *Bull. Eng. Geol. Environ.* **2019**, *78*, 5977–5989. [CrossRef]
28. Martha, T.R.; Kerle, N.; Jetten, V.; van Westen, C.J.; Kumar, K.V. Landslide Volumetric Analysis Using Cartosat-1-Derived DEMs. *Ieee Geosci. Remote Sens. Lett.* **2010**, *7*, 582–586. [CrossRef]
29. Cruden, D.M.; Varnes, D.J. Landslides: Investigation and mitigation. In *National Research Council Transportation Research Board Special Report (Book 247)*; Turner, A.K., Schuster, R.L., Eds.; Transportation Research Board: Washington, DC, USA, 1996; pp. 36–75.
30. Vinod kumar, K.; Lakhera, R.C.; Martha, T.R.; Chatterjee, R.S.; Bhattacharya, A. Analysis of the 2003 Varunawat Landslide, Uttarkashi, India using Earth Observation data. *Environ. Geol.* **2008**, *55*, 789–799. [CrossRef]
31. Martha, T.R.; Kerle, N.; van Westen, C.J.; Jetten, V.; Kumar, K.V. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. *Isprs J. Photogramm. Remote Sens.* **2012**, *67*, 105–119. [CrossRef]
32. Martha, T.R.; Kamala, P.; Jose, J.; Kumar, K.V.; Sankar, G.J. Identification of new Landslides from High Resolution Satellite Data Covering a Large Area Using Object-Based Change Detection Methods. *J. Indian Soc. Remote Sens.* **2016**, *44*, 515–524. [CrossRef]
33. Amatya, P.; Kirschbaum, D.; Stanley, T. Use of Very High-Resolution Optical Data for Landslide Mapping and Susceptibility Analysis along the Karnali Highway, Nepal. *Remote Sens.* **2019**, *11*, 2284. [CrossRef]
34. Vamsee, A.M.; Kamala, P.; Martha, T.R.; Kumar, K.V.; Amminedu, E. A tool assessing optimal multi-scale image segmentation. *J. Indian Soc. Remote Sens.* **2018**, *46*, 31–41. [CrossRef]
35. Martha, T.R.; Kerle, N.; Jetten, V.; van Westen, C.J.; Kumar, K.V. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods. *Geomorphology* **2010**, *116*, 24–36. [CrossRef]
36. Martha, T.R.; van Westen, C.J.; Kerle, N.; Jetten, V.; Kumar, K.V. Landslide hazard and risk assessment using semi-automatically created landslide inventories. *Geomorphology* **2013**, *184*, 139–150. [CrossRef]
37. Kumar, V.; Gupta, V.; Sundriyal, Y.P. Spatial interrelationship of landslides, litho-tectonics, and climate regime, Satluj valley, Northwest Himalaya. *Geol. J.* **2019**, *54*, 537–551. [CrossRef]
38. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.-T. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. [CrossRef]
39. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196. [CrossRef]
40. Guzzetti, F.; Peruccacci, S.; Rossi, M.; Stark, C.P. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol. Atmos. Phys.* **2007**, *98*, 239–267. [CrossRef]
41. Dikshit, A.; Sarkar, R.; Pradhan, B.; Acharya, S.; Dorji, K. Estimating rainfall thresholds for landslide occurrence in the Bhutan Himalayas. *Water* **2019**, *11*, 1616. [CrossRef]
42. Dikshit, A.; Sarkar, R.; Pradhan, B.; Jena, R.; Drukpa, D.; Alamri, A.M. Temporal Probability Assessment and Its Use in Landslide Susceptibility Mapping for Eastern Bhutan. *Water* **2020**, *12*, 267. [CrossRef]

43. Segoni, S.; Piciullo, L.; Gariano, S.L. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **2018**, *15*, 1483–1501. [CrossRef]
44. Gariano, S.L.; Sarkar, R.; Dikshit, A.; Dorji, K.; Brunetti, M.T.; Peruccacci, S.; Melillo, M. Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. *Bull. Eng. Geol. Environ.* **2019**, *78*, 4325–4332. [CrossRef]
45. Sengupta, A.; Gupta, S.; Anbarasu, K. Rainfall thresholds for the initiation of landslide at Lanta Khola in north Sikkim, India. *Nat. Hazards* **2010**, *52*, 31–42. [CrossRef]
46. Dikshit, A.; Sarkar, R.; Satyam, N. Probabilistic approach toward Darjeeling Himalayas landslides-A case study. *Cogent Eng.* **2018**, 5. [CrossRef]
47. Teja, T.S.; Dikshit, A.; Satyam, N. Determination of Rainfall Thresholds for Landslide Prediction Using an Algorithm-Based Approach: Case Study in the Darjeeling Himalayas, India. *Geosciences* **2019**, 302. [CrossRef]
48. Harilal, G.T.; Madhu, D.; Ramesh, M.V.; Pullarkatt, D. Towards establishing rainfall thresholds for a real-time landslide early warning system in Sikkim, India. *Landslides* **2019**, *16*, 2395–2408. [CrossRef]
49. Gariano, S.L.; Melillo, M.; Peruccacci, S.; Brunetti, M.T. How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering? *Nat. Hazards* **2019**, 1–16. [CrossRef]
50. Mathew, J.; Babu, D.G.; Kundu, S.; Kumar, K.V.; Pant, C.C. Integrating intensity-duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides* **2014**, *11*, 575–588. [CrossRef]
51. Kumar, A.; Asthana, A.L.; Priyanka, R.S.; Jayangondaperumal, R.; Gupta, A.K.; Bhakuni, S.S. Assessment of landslide hazards induced by extreme rainfall event in Jammu and Kashmir Himalaya, northwest India. *Geomorphology* **2017**, *284*, 72–87. [CrossRef]
52. Caine, N. The rainfall intensity-duration control of shallow landslides and debris flows. *Geogr. Ann. Ser. APhys. Geogr.* **1980**, *62*, 23–27.
53. Chae, B.-G.; Park, H.-J.; Catani, F.; Simoni, A.; Berti, M. Landslide prediction, monitoring and early warning: A concise review of state-of-the-art. *Geosci. J.* **2017**, *21*, 1033–1070. [CrossRef]
54. Dikshit, A.; Satyam, D.N.; Towhata, I. Early warning system using tilt sensors in Chibo, Kalimpong, Darjeeling Himalayas, India. *Nat. Hazards* **2018**, *94*, 727–741. [CrossRef]
55. Falae, P.O.; Kanungo, D.P.; Chauhan, P.K.S.; Dash, R.K. Electrical resistivity tomography (ERT) based subsurface characterisation of Pakhi Landslide, Garhwal Himalayas, India. *Environ. Earth Sci.* **2019**, 78. [CrossRef]
56. Yhokha, A.; Goswami, P.K.; Chang, C.P.; Yen, J.Y.; Ching, K.E.; Aruche, K.M. Application of Persistent Scatterer Interferometry (PSI) in monitoring slope movements in Nainital, Uttarakhand Lesser Himalaya, India. *J. Earth Syst. Sci.* **2018**, 127. [CrossRef]
57. Martha, T.R.; Reddy, P.S.; Bhatt, C.M.; Raj, K.B.G.; Nalini, J.; Padmanabha, E.A.; Narender, B.; Kumar, K.V.; Muralikrishnan, S.; Rao, G.S.; et al. Debris volume estimation and monitoring of Phuktal river landslide-dammed lake in the Zanskar Himalayas, India using Cartosat-2 images. *Landslides* **2017**, *14*, 373–383. [CrossRef]
58. Mondal, S.K.; Sastry, R.G.; Pachauri, A.K.; Gautam, P.K. High resolution 2D electrical resistivity tomography to characterize active Naitwar Bazar landslide, Garhwal Himalaya, India. *Curr. Sci.* **2008**, *94*, 871–875.
59. Kannaujiya, S.; Chattoraj, S.L.; Jayalath, D.; Ray, P.K.C.; Bajaj, K.; Podali, S.; Bisht, M.P.S. Integration of satellite remote sensing and geophysical techniques (electrical resistivity tomography and ground penetrating radar) for landslide characterization at Kunjethi (Kalimath), Garhwal Himalaya, India. *Nat. Hazards* **2019**, *97*, 1191–1208. [CrossRef]
60. Sharma, S.P.; Anbarasu, K.; Gupta, S.; Sengupta, A. Integrated very low-frequency EM, electrical resistivity, and geological studies on the Lanta Khola landslide, North Sikkim, India. *Landslides* **2010**, *7*, 43–53. [CrossRef]
61. Raj, K.; Govindha, B.; Martha, T.R.; Kumar, K.V. A bird's-eye view of landslide dammed lakes in Zanskar Himalaya, India. *Curr. Sci.* **2017**, *112*, 1109–1112.
62. Weidinger, J.T. Case history and hazard analysis of two lake-damming landslides in the Himalayas. *J. Asian Earth Sci.* **1998**, *16*, 323–331. [CrossRef]
63. Kumar, V.; Gupta, V.; Jamir, I.; Chattoraj, S.L. Evaluation of potential landslide damming: Case study of Urni landslide, Kinnaur, Satluj valley, India. *Geosci. Front.* **2019**, *10*, 753–767. [CrossRef]

64. Gupta, V.; Bhasin, R.K.; Kaynia, A.M.; Kumar, V.; Saini, A.S.; Tandon, R.S.; Pabst, T. Finite element analysis of failed slope by shear strength reduction technique: A case study for Surabhi Resort Landslide, Mussoorie township, Garhwal Himalaya. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1677–1690. [CrossRef]
65. Gupta, V.; Sah, M. Impact of the trans-Himalayan landslide lake outburst flood (LLOF) in the Satluj catchment, Himachal Pradesh, India. *Nat. Hazards* **2008**, *45*, 379–390. [CrossRef]
66. Gariano, S.L.; Guzzetti, F. Landslides in a changing climate. *Earth-Sci. Rev.* **2016**, *162*, 227–252. [CrossRef]
67. Sarkar, S.; Kanungo, D. Landslide Disaster on Berinag-Munsiyari Road, Pithoragarh District, Uttarakhand. *Curr. Sci.* **2010**, *98*, 2.
68. Bhan, S.; Devrani, A.; Sinha, V. An analysis of monthly rainfall and the meteorological conditions associated with cloudburst over the dry region of Leh (Ladakh), India. *Mausam* **2015**, *66*, 107–122.
69. Anbalagan, R. Landslide hazard evaluation and zonation mapping in mountainous terrain. *Eng. Geol.* **1992**, *32*, 269–277. [CrossRef]
70. Gupta, P.; Anbalagan, R. Slope stability of Tehri Dam Reservoir Area, India, using landslide hazard zonation (LHZ) mapping. *Q. J. Eng. Geol.* **1997**, *30*, 27–36. [CrossRef]
71. Kanungo, D.; Arora, M.; Gupta, R.; Sarkar, S. Landslide risk assessment using concepts of danger pixels and fuzzy set theory in Darjeeling Himalayas. *Landslides* **2008**, *5*, 407–416. [CrossRef]
72. Sarkar, S.; Roy, A.K.; Raha, P. Deterministic approach for susceptibility assessment of shallow debris slide in the Darjeeling Himalayas, India. *Catena* **2016**, *142*, 36–46. [CrossRef]
73. Mathew, J.; Kundu, S.; Kumar, K.V.; Pant, C.C. Hydrologically complemented deterministic slope stability analysis in part of Indian Lesser Himalaya. *Geomat. Nat. Hazards Risk* **2016**, *7*, 1557–1576. [CrossRef]
74. Kumar, R.; Anbalagan, R. Landslide susceptibility zonation in part of Tehri reservoir region using frequency ratio, fuzzy logic and GIS. *J. Earth Syst. Sci.* **2015**, *124*, 431–448. [CrossRef]
75. Ghosh, S.; Carranza, E.J.M.; van Westen, C.J.; Jetten, V.G.; Bhattacharya, D.N. Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). *Geomorphology* **2011**, *131*, 35–56. [CrossRef]
76. Mandal, S.; Mandal, K. Modeling and mapping landslide susceptibility zones using GIS based multivariate binary logistic regression (LR) model in the Rorachu river basin of eastern Sikkim Himalaya, India. *Modeling Earth Syst. Environ.* **2018**, *4*, 69–88. [CrossRef]
77. Mathew, J.; Jha, V.K.; Rawat, G.S. Application of binary logistic regression analysis and its validation for landslide susceptibility mapping in part of Garhwal Himalaya, India. *Int. J. Remote Sens.* **2007**, *28*, 2257–2275. [CrossRef]
78. Pourghasemi, H.R.; Yansari, Z.T.; Panagos, P.; Pradhan, B. Analysis and evaluation of landslide susceptibility: A review on articles published during 2005–2016 (periods of 2005–2012 and 2013–2016). *Arab. J. Geosci.* **2018**, *11*, 193. [CrossRef]
79. Sharma, S.; Mahajan, A.K. A comparative assessment of information value, frequency ratio and analytical hierarchy process models for landslide susceptibility mapping of a Himalayan watershed, India. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2431–2448. [CrossRef]
80. Das, I.; Stein, A.; Kerle, N.; Dadhwal, V.K. Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. *Geomorphology* **2012**, *179*, 116–125. [CrossRef]
81. Mathew, J.; Jha, V.K.; Rawat, G.S. Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method. *Landslides* **2009**, *6*, 17–26. [CrossRef]
82. Chauhan, S.; Mukta, S.; Arora, M.K. Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model. *Landslides* **2010**, *7*, 411–423. [CrossRef]
83. Sahana, M.; Sajjad, H. Evaluating effectiveness of frequency ratio, fuzzy logic and logistic regression models in assessing landslide susceptibility: A case from Rudraprayag district, India. *J. Mt. Sci.* **2017**, *14*, 2150–2167. [CrossRef]
84. Saha, A.K.; Gupta, R.P.; Sarkar, I.; Arora, M.K.; Csaplovics, E. An approach for GIS-based statistical landslide susceptibility zonation - with a case study in the Himalayas. *Landslides* **2005**, *2*, 61–69. [CrossRef]
85. Sarkar, S.; Anbalagan, R. Landslide hazard zonation mapping and comparative analysis of hazard zonation maps. *J. Mt. Sci.* **2008**, *5*, 232–240. [CrossRef]

86. Mondal, S.; Maiti, R. Integrating the Analytical Hierarchy Process (AHP) and the Frequency Ratio (FR) Model in Landslide Susceptibility Mapping of Shiv-khola Watershed, Darjeeling Himalaya. *Int. J. Disaster Risk Sci.* **2013**, *4*, 200–212. [CrossRef]
87. Sharma, L.P.; Patel, N.; Ghose, M.K.; Debnath, P. Application of frequency ratio and likelihood ratio model for geo-spatial modelling of landslide hazard vulnerability assessment and zonation: A case study from the Sikkim Himalayas in India. *Geocarto Int.* **2014**, *29*, 128–146. [CrossRef]
88. Balamurugan, G.; Ramesh, V.; Touthang, M. Landslide susceptibility zonation mapping using frequency ratio and fuzzy gamma operator models in part of NH-39, Manipur, India. *Nat. Hazards* **2016**, *84*, 465–488. [CrossRef]
89. Manzo, G.; Tofani, V.; Segoni, S.; Battistini, A.; Catani, F. GIS techniques for regional-scale landslide susceptibility assessment: The Sicily (Italy) case study. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1433–1452. [CrossRef]
90. Singh, C.; Kohli, A.; Kumar, P. Comparison of results of BIS and GSI guidelines on macrolevel landslide hazard zonation—A case study along highway from Bhalukpong to Bomdila, West Kameng district, Arunachal Pradesh. *J. Geol. Soc. India* **2014**, *83*, 688–696. [CrossRef]
91. Ghosh, S.; van Westen, C.J.; Carranza, E.J.M.; Ghoshal, T.B.; Sarkar, N.K.; Surendranath, M. A quantitative approach for improving the BIS (Indian) method of medium-scale landslide susceptibility. *J. Geol. Soc. India* **2009**, *74*, 625–638. [CrossRef]
92. Das, I.; Stein, A.; Kerle, N.; Dadhwal, V.K. Probabilistic landslide hazard assessment using homogeneous susceptible units (HSU) along a national highway corridor in the northern Himalayas, India. *Landslides* **2011**, *8*, 293–308. [CrossRef]
93. Sarkar, S.; Kanungo, D.P.; Sharma, S. Landslide hazard assessment in the upper Alaknanda valley of Indian Himalayas. *Geomat. Nat. Hazards Risk* **2015**, *6*, 308–325. [CrossRef]
94. Pham, B.T.; Bui, D.T.; Prakash, I.; Dholakia, M.B. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena* **2017**, *149*, 52–63. [CrossRef]
95. Pham, B.T.; Prakash, I.; Dou, J.; Singh, S.K.; Trinh, P.T.; Tran, H.T.; Le, T.M.; Phong, T.V.; Khoi, D.K.; Shirzadi, A.; et al. A novel hybrid approach of landslide susceptibility modelling using rotation forest ensemble and different base classifiers. *Geocarto Int.* **2019**. [CrossRef]
96. Peethambaran, B.; Anbalagan, R.; Shihabudheen, K.V.; Goswami, A. Robustness evaluation of fuzzy expert system and extreme learning machine for geographic information system-based landslide susceptibility zonation: A case study from Indian Himalaya. *Environ. Earth Sci.* **2019**, 78. [CrossRef]
97. Ramakrishnan, D.; Singh, T.N.; Verma, A.K.; Gulati, A.; Tiwari, K.C. Soft computing and GIS for landslide susceptibility assessment in Tawaghat area, Kumaon Himalaya, India. *Nat. Hazards* **2013**, *65*, 315–330. [CrossRef]
98. Pham, B.T.; Bui, D.T.; Prakash, I.; Nguyen, L.H.; Dholakia, M.B. A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environ. Earth Sci.* **2017**, 76. [CrossRef]
99. Pham, B.T.; Bui, D.T.; Pourghasemi, H.R.; Indra, P.; Dholakia, M. Landslide susceptibility assesssment in the Uttarakhand area (India) using GIS: A comparison study of prediction capability of naïve bayes, multilayer perceptron neural networks, and functional trees methods. *Theor. Appl. Climatol.* **2017**, *128*, 255–273. [CrossRef]
100. Pham, B.T.; Prakash, I.; Singh, S.K.; Shirzadi, A.; Shahabi, H.; Tran, T.T.G.; Buig, D.T. Landslide susceptibility modeling using Reduced Error Pruning Trees and different ensemble techniques: Hybrid machine learning approaches. *Catena* **2019**, *175*, 203–218. [CrossRef]
101. Chauhan, S.; Sharma, M.; Arora, M.K.; Gupta, N.K. Landslide Susceptibility Zonation through ratings derived from Artificial Neural Network. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, 340–350. [CrossRef]
102. Kumar, D.; Thakur, M.; Dubey, C.S.; Shukla, D.P. Landslide susceptibility mapping & prediction using Support Vector Machine for Mandakini River Basin, Garhwal Himalaya, India. *Geomorphology* **2017**, *295*, 115–125. [CrossRef]
103. Pham, B.T.; Shirzadi, A.; Bui, D.T.; Prakash, I.; Dholakia, M. A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: A case study in the Himalayan area, India. *Int. J. Sediment Res.* **2018**, *33*, 157–170. [CrossRef]

104. Pham, B.T.; Prakash, I.; Jaafari, A.; Bui, D.T. Spatial prediction of rainfall-induced landslides using aggregating one-dependence estimators classifier. *J. Indian Soc. Remote Sens.* **2018**, *46*, 1457–1470. [CrossRef]
105. Pham, B.T.; Jaafari, A.; Prakash, I.; Bui, D.T. A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bull. Eng. Geol. Environ.* **2019**, *78*, 2865–2886. [CrossRef]
106. Jaafari, A.; Panahi, M.; Pham, B.T.; Shahabi, H.; Bui, D.T.; Rezaie, F.; Lee, S. Meta optimization of an adaptive neuro-fuzzy inference system with grey wolf optimizer and biogeography-based optimization algorithms for spatial prediction of landslide susceptibility. *Catena* **2019**, *175*, 430–445. [CrossRef]
107. Kanungo, D.P.; Arora, M.K.; Sarkar, S.; Gupta, R.P. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng. Geol.* **2006**, *85*, 347–366. [CrossRef]
108. Kanungo, D.P.; Sarkar, S.; Sharma, S. Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. *Nat. Hazards* **2011**, *59*, 1491–1512. [CrossRef]
109. Chawla, A.; Chawla, S.; Pasupuleti, S.; Rao, A.C.S.; Sarkar, K.; Dwivedi, R. Landslide Susceptibility Mapping in Darjeeling Himalayas, India. *Adv. Civ. Eng.* **2018**. [CrossRef]
110. Meena, S.R.; Mishra, B.K.; Piralilou, S.T. A Hybrid Spatial Multi-Criteria Evaluation Method for Mapping Landslide Susceptible Areas in Kullu Valley, Himalayas. *Geosciences* **2019**, 156. [CrossRef]
111. Tofani, V.; Bicocchi, G.; Rossi, G.; Segoni, S.; D'Ambrosio, M.; Casagli, N.; Catani, F. Soil characterization for shallow landslides modeling: A case study in the Northern Apennines (Central Italy). *Landslides* **2017**, *14*, 755–770. [CrossRef]
112. Segoni, S.; Tofani, V.; Rosi, A.; Catani, F.; Casagli, N. Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Front. Earth Sci.* **2018**, *6*, 85. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-08655-6

Real-time visual tracking via multi-cue based adaptive particle filter framework

Ashish Kumar[1] · **Gurjit Singh Walia**[2] · **Kapil Sharma**[1]

Received: 6 January 2019 / Revised: 2 December 2019 / Accepted: 3 January 2020 /
Published online: 22 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Visual tracking using particle filter has been extensively investigated due to its myriad of application in the field of computer vision. However, particle filter framework performance is heavily impaired due to its inherent problems namely, particle degeneracy and impoverishment. In addition, most of the tracking methods using single cue are greatly affected by dynamic environmental challenges. To address these issues, we propose an adaptive multi-cue particle filter based real-time visual tracking framework. Three complementary cues namely, color histogram, LBP and pyramid of histogram of gradient have been exploited for object's appearance model. These cues are integrated using the proposed adaptive fusion model for the automatic boosting of important particles and suppression of unimportant particles. Resampling method using butterfly search optimization relocate low performing particles to high likelihood area. Proposed outlier detection mechanism not only helps in detecting low performing particles but also aids in updating of the reference dictionary. Online estimation of cue reliability along with its multi-cue fusion leads to quick adaptation of the proposed tracker. On average of the outcome, our tracker achieves average center location error of 6.89 (in pixels) and average F-measure of 0.786 when evaluated on OTB-100 and VOT dataset against 13 others state-of-the-art.

Keywords Particle filter · Visual tracking · Fusion model · Cue reliability · Resampling

✉ Kapil Sharma
kapil@ieee.org

Ashish Kumar
ashish.gupta14d@gmail.com

Gurjit Singh Walia
gurjit.walia@gmail.com

[1] Delhi Technological University, Delhi, India

[2] SAG, Defense Research and Development Organization, New Delhi, India

1 Introduction

Object tracking in video sequences is one of the active research areas in computer vision. The objective of visual tracking is to localize the target during distinct frames in a video sequence. However, target localization in subsequent frames is challenging due to complex background clutters, variations in scale and pose, full or partial occlusion, noise in the image and real-time constraints. Various visual tracking frameworks were explored and discussed at length [22]. Under stochastic object tracking framework, the non-linear and non-Gaussian tracking model is best presented as particle filter based tracking.

Particle filter (PF) was defined as bootstrap filter [14] and used for state estimation using Sequential Monte Carlo methods. It consists of two steps: 1) Prediction, which evolves particles using state model 2) likelihood calculation, which determines particles weight during update state. The main advantage of PF is that it reduces the batch of sample patches during tracking and handle the dynamic environment challenges. Due to this, PF has been widely used in many real-time tracking frameworks [38, 39]. However, the performance of PF is restricted by its inherent shortcoming of sample degeneracy. Sample degeneracy is the problem when most of the particles weight become so negligible that they do not contribute much towards state estimation. Many resampling techniques [8, 20] were studied to address the PF drawback. Generally, resampling techniques replace the low weight particles but this process reduce the diversity in the search space. Hence, most of the resampling methods leads to sample impoverishment where all the particles accumulated to small area. To solve this, many nature-based optimization viz. Firefly algorithm [13], Improved cuckoo search [31], Ant optimization [35], Particle Swarm Optimization (PSO) [42] were explored to a great extent. These methods improve PF based trackers efficiency by addressing its problems. These techniques were applied as resampling methods in particle filter tracking framework in which the appearance model is constructed either through single cue or multi-cue.

Generally, single cue viz. color, texture, intensity, gradient, etc. were used in PF framework [13, 31, 35]. It has been well argued that single cue is not sufficient to cater dynamic environmental challenges [32]. Recently, a lot of work have been investigated under multi-cue object tracking [30]. For robust object tracking, complementary cues information needs to be considered so that if one cue fails then other may compensate. In [38], author exploited color and motion information of the target in a PF based tracking framework. Likelihood of each cue is multiplied with adaptive reliability and then, summed to generate the fused likelihood. Similarly, Wen et al. [36] used color and motion information of the target but integrate them using linear weighted method. Firouznia et al. [12] used motion information to estimate the target's position globally and then the color information was extracted to localize target in the local region under chaotic PF framework. In [33], author exploited color, thermal and motion cue for observation model. Conflict amongst various cue was resolved using Dsmt based PCR-5 rules. Dou and Li [9] used Corrected Background-Weighted Histogram (CBWH), Completed Local Ternary Patterns (CLTP) and Histogram of Oriented Gradients (HOG) for the target model. Multiple interactive model probabilities were used for particle importance. In [21], authors had extracted color and intensity features to localize the target in visual attention integrated PF. Salient regions were searched to detect the target in the video but the proposed work was inefficient in handling target's appearance variations due to changing environment. In [10], color, texture and HOG cue were determined for each particle. Linear normalized weighted fusion of cues was proposed by integrating reliability at each frame. Linear fusion of cues was less efficient in capturing the discriminative capability of each exploited cue. In addition, the proposed method

was computationally slow and inefficient in handling the background clutters. Xiao et al. [38] exploited color and shape feature of the target and fused them by normalized weighted feature fusion. Model was made adaptive using the contextual information. Contextual information for each cue was determined by analyzing the particles in the search space based on the foreground and background information already encoded in them. However, experimental results revealed that there was no significant improvement in many video sequences by the adaptive fusion of feature saliency. Walia et al. [34] integrated color and LBP features using adaptive fuzzy based fusion model. The method had shown real-time performance but inefficient in handling the fast motion and motion blur challenges. Zhang et al. [44] proposed a tracking algorithm exploiting color, HOG and CNN features. Interdependencies between each feature was exploited using incremental strategy in the update model. The method had addressed the particle filter limitations to a great extent but was computationally expensive and not suitable for real-time tracking problems. In sum, the existing works have limitedly addressed real-time tracking problem which can handle various environmental variations namely, illumination variation, background clutters, full or partial occlusion, scale variations, deformation, fast motion and motion blur. The purpose of this research is to propose a novel adaptive multi-cue tracking model in PF framework which can not only address environmental challenges but also suitable for real-time tracking. The details of the proposed work is follow in turn.

In this paper, a multi-cue object tracking method has been proposed under PF framework. Issues of particle filter are addressed by exploiting a resampling method based on Butterfly optimization algorithm. A novel adaptive fusion of multi-cue is also proposed to boost and suppress the particles automatically during occlusion and background clutters. At each stage, context sensitive cue reliability is estimated for the quick adaptation of the tracker under illumination and scale variations. In addition, state model also caters for the change in scale and rotation of the target. In order to improve the computational efficiency of the proposed method, an outlier detection mechanism is embedded to identify the low performing particles. The key contributions of the work are summarized as follows:

- We propose a tracker under particle filter framework in which likelihood model is constructed by considering three complementary multi-cues namely color histogram, linear binary pattern (LBP) and Pyramid Histograms of oriented gradients (PHOG). RGB cue is robust to change in scale and partial occlusion but deteriorates during illumination variations and background clutter. However, LBP texture cue is invariant to change in illumination and background clutters but cannot handle scale variations and target's deformation. while the PHOG can efficiently handle the target's deformation and rotation. These complementary cues are considered with the aim so that if one cue fail the other may compensate during the dynamic environment.
- Resampling technique based on Butterfly search optimization is proposed to address the shortcomings of PF. This optimization has a fast convergence rate with its two parameters: 1) Switch probability 2) Solution vector. These two parameters diversify the particles in high likelihood area of search space to prevent particle degeneracy and sample impoverishment.
- An adaptive fusion method for fusion of multi-cue is proposed to cater dynamic environmental variations. This fusion model ensures the automatic boosting of the important particles and suppression of the unimportant particles during occlusion, background clutter and target's deformation. In addition, Context sensitive cue reliability is estimated to ensure the quick adaptation of the tracker during real-time environmental challenges.

– An outlier detection mechanism is also proposed to detect the low performing particles as outliers. In each subsequent frame, outliers are detected which dispense less towards the state estimation. This mechanism not only helps in improving the efficiency of the proposed resampling method but also reduces the computational complexity of the proposed tracker.

The remainder of this paper is organized as follow. Related works closely in the domain of the proposed work are revived in Section 2. Section 3 presents the core architecture of the proposed tracker. To prove the robustness and efficiency of the proposed method, experimental results are discussed in Section 4. Also, performance of the proposed tracker is compared with other trackers on benchmarked data sets. Finally, the conclusion and future directions are sketched in Section 5.

2 Related work: particle filter based visual tracking

In this section, we have reviewed the closely related literature to our work. Visual tracking under particle filter framework has been investigated by many researchers using either single cue or multi-cue. Single cue namely color, texture, shape, motion or gradient has its own limitation which can be addressed by using multi-cue in PF framework. In addition, the tracker's efficiency under PF framework is limited by its shortcoming of particle degeneracy and impoverishment. In sum, there is scope of improvement of trackers performance in multi-cue PF framework. Resampling techniques and multi-cue fusion in PF framework require further investigation for efficient tracking solutions.

In order to address the dynamic environment multi-cue based PF framework were investigated. In this direction, Walia et al. [33] considered color, motion and thermal information about the target in PF framework. Multi-cues were integrated by resolving conflicts by PCR-5 rule. It supported automatic boosting and suppression of the cues to assist cue fusion adaptively. In [10], authors considered color histogram, texture and HOG as features descriptor and calculated the predicted reliability for their adaptive linear fusion. Xiao et al. [38] proposed adaptive data fusion for color and shape cues by continuous relearning model to cater changing appearance of the target during tracking. Two-step estimation was exploited to detect and replace drifting particles. In [11], authors exploited color information for a color image alongwith intensity and gradient for a grey image in an iterative PF framework. Huan et al. [16] proposed a tracking solution considering color and LBP feature to handle occlusion. However, authors [23] has exploited colour and motion information for occlusion handling using SURF features. But real time application for these method was not investigated. Similarly, In [25], authors used colour and motion cue in a pre-trained Convolutional neural network (CNN) under PF framework. However, embedding pre-train CNN in particle filter was computationally expensive. In sum, multi-cue under PF framework requires further improvement to provide robust tracking solutions.

In addition, the performance of PF framework was improved through resampling methods based on meta-heuristic approaches. In this direction, authors proposed modified galaxy-based search algorithm as resampling technique for improving PF performance [27]. Color, HOG and motion cues were exploited but the proposed method was inefficient in handling occlusion. Zhang et al. [42] exploited PSO as resampling technique and determined intensity and gradient cues for each particle. Discriminative weight of each cue was fused adaptively at each frame. In [15], authors proposed immune genetic algorithm as resampling method to address PF problems. Crossover, mutation, promotion and inhibition

parameters were used to ensure particle diversity in search space. However, the performance of this method was evaluated only on two video sequences. On the other hand, Yin et al. [39] used color and motion cues in particle filter based object tracking. CamShift exploited as optimization to determine the local maxima during environmental variations. However, the method was evaluated on limited performance metrics. In [31], authors proposed improved cuckoo search to solve the problems of PF for object tracking. Simulation results on video sequences were presented to prove the robustness of the proposed methods. In [36], author extracted two features from the target and fused them using linear weighted method. In sum, resampling methods under multi-cue particle filter framework for tracking the object in video sequence are investigated. However, the shortcomings of particle filter with adaptive fusion methodology are still need to be addressed for efficient object tracking. Details about the architecture of the proposed real-time adaptive tracker has been discussed in the next section.

3 Proposed tracker architecture

The proposed method utilizes complementary cues under multi-cue PF framework for tracking the object in video sequences. Adaptive update model with context sensitive cue reliability has been proposed for tracker's quick adaptation to environmental variations. In addition, problems of PF are addressed by proposing butterfly search optimization based resampling method. Architecture of the proposed tracker, update model and the proposed resampling technique are depicted in Fig. 1. Initially, the target is segmented from the background using the GMM subtraction [29]. Particles (N) are instantiated around the centroid

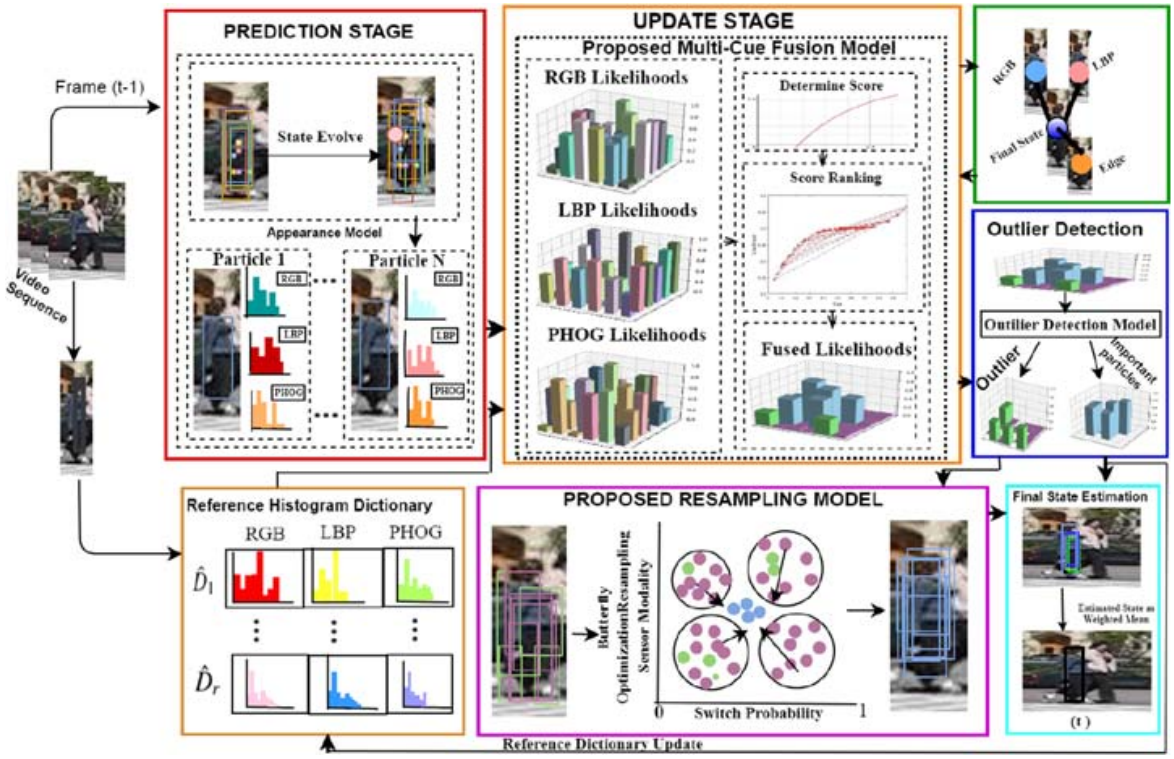


Fig. 1 Architecture of the proposed approach. At each time step t, the particles are initialized and are evolved through the state model. Each particle is evaluated for three cues and reference histogram dictionary is extracted. Cues are integrated through an adaptive fusion model with context sensitive cue reliability and outlier detection. Particles are resampled and final state is estimated as the weighted mean of the outliers and important particles

of the detected target. All of these particles are evolved through the state vector (S_t) defined by the multi-component state model [7]. Predicted particle is evaluated for each features descriptor namely color histogram (F_i), texture (F_e) and edge (F_g) individually. These cue are further subjected to adaptive fusion approach to obtain the fused weight ($\hat{W}_f$). This model ensures the automatic boosting and suppression of the particles during environmental challenges. At each time step t, context sensitive cue reliability (r_t^l) is also calculated to discount the particles based on each cue performance. The particles so obtained are passed through an outlier detection mechanism to detect the low performing particles ($\hat{U}_x$) as outliers and to detect important particles ($\hat{I}_y$). Outliers are passed through a butterfly optimization based resampling technique to obtain (U_x). This resampling technique diversifies the unimportant particles in the search area with its two variables viz. sensor modality and switch probability. Sensor modality helps in sensing the location of the particles and switch probability propagates the particles in the high likelihood region. Final state (G_t) of the target is estimated by the weighted sum of the resampled outliers (U_x) and the important particles ($\hat{I}_y$). This process is repeated iteratively for the video sequence. Reference dictionary ($\hat{D}_r$) is updated by selective replacement of the important particles. This ensures consistent updating of the appearance model at each time step in accordance with environmental variations. Core design of the proposed approach is as follows.

3.1 Core design of multi-cue based adaptive tracker

In this section, core design of the multi-cue based adaptive tracker is discussed. The details of the multi-cue particle filter alongwith multi-cue feature extraction and adaptive multi-cue fusion model are as follow.

3.1.1 Multi-cue particle filter

Particle filter utilizes Monte Carlo algorithm for solving the state estimation problems. Particle filter has two stages: prediction stage and update stage to obtain the desired PDF. During the prediction stage, particle state is evolved through the system model at time t using (1).

$$S_t = f_{t-1}(S_{t-1}, \sigma_{t-1}), \tag{1}$$

where, f_{t-1} is non-linear state transition function and σ_{t-1} is zero mean white noise. Particles are distributed on the target to estimate the posterior density distribution $p(S_t|x_t)$ where, $x_t = x_1, x_2...x_t$ is the set of available information at time t. The measurement x_t at time t is determined by the observation model using (2).

$$x_t = w_t(S_t, K_t) \tag{2}$$

where, w_t is the non-linear function and K_t is zero mean white noise, independent of present and past state. If PDF $p(S_{t-1}|x_{t-1})$ is available at time $t-1$, then the prior PDF is obtained using the (3).

$$p(S_t|x_{t-1}) = \int p(S_t|S_{t-1})p(S_{t-1}|x_{t-1})dS_{t-1}, \tag{3}$$

Finally, when the measurement x_t is available at time t, then the state is updated using Baye's rule given by (4).

$$p(S_t|x_t) = \frac{p(x_t|S_t)p(S_t|x_{t-1})}{p(x_t|x_{t-1})} \tag{4}$$

Multi-component model [7] defines state of the particle by the state vector S_t which includes variables namely X_p, Y_p, V_{xp}, V_{yp}, ro_p and α_p. (X_p, Y_p) is the center of rectangle of bounding box. V_{xp} and V_{yp} are velocities in respective directions for p_{th} particle modeled through constant velocity model. ro_p and α_p are rotation and scaling variables of the particle which used random walk model.

Predicted particles are obtained by evolving the particles through the state model using (1). Each predicted particle is subjected to multi-cue evaluation namely color histogram (F_i), texture (F_e) and edge (F_g). Color cue is stable during change in scale and partial occlusion, texture cue can handle illumination variations and background clutters, and edge cue is invariant object's deformation and rotation. These are complementary cues i.e. if one fade others may compensate during the tracking challenges. Details of the likelihood calculation for these cues is discussed in the following section.

3.1.2 Multi-cue feature extraction

The proposed tracking approach considers three complementary cues namely, color, LBP and PHOG for each predicted particles. The details of each cue alongwith their likelihood calculation is as follows:

RGB color model is exploited for calculating the color histogram. Color histogram for p_{th} particle is $F_i^p = H_1, H_2...H_{\hat{N}_b}$ and determined using (5).

$$H_c = Z \sum_{q=1}^{\hat{N}} B(m_q, n_q), c = 1, 2....\hat{N}_b \tag{5}$$

where Z is the normalizing constant and $B(.)$ represents binning function that assign pixel (m_q, n_q) to one of the $\hat{N}_b$ histogram bins. For each particle, color histogram (F_i^p) is calculated by mapping it to normalized H_c.

Texture cue measures intensity variations of the targets surface [24]. It is calculated for each particle using scale-invariant LBP using (6).

$$LBP_{\alpha,\hat{N}_\eta} = \begin{cases} \sum_{r=0}^{\hat{N}_\eta - 1} (L(G^k - G^c)), & \text{if} U(LBP_{\alpha,\hat{N}_\eta}) \leq 2 \\ \eta + 1, & \text{otherwise} \end{cases} \tag{6}$$

where, G^k is grey value for k_{th} pixel and G^c is grey value for central pixel(c) with radius α and equally spaced neighboring pixels $\hat{N}_\eta$. $L(.)$ is the step function. $U(LBP_{\alpha,\hat{N}_\eta})$ determines the uniform pattern in the image of the pattern labels. For each particle, LBP histogram (F_e) is determined by normalizing the value of the $LBP_{\alpha,\hat{N}_\eta}$.

Spatial distribution of edges is considered for the representation of target's shape. PHOG (Pyramid of Histogram of gradients) [6] was used for the extraction of shape information of the target with its spatial distribution. In this work, PHOG is extracted for each particle considering the target's shape and edge orientation. For this, Region Of Interest $(\mathcal{I})$ is extracted and then intensity gradient values are computed for pixel (m_q, n_q) using (7).

$$E(m_q, n_q) = \sqrt{(\mathcal{I}(m_q, n_{q+1}) - \mathcal{I}(m_q, b_{q-1}))^2 + (\mathcal{I}(m_{q+1}, n_q) - \mathcal{I}(m_{q-1}, n_q))^2} \tag{7}$$

For a given image, 20 bins are utilized using orientation which is depicted using (8)

$$\theta = \tan^{-1}\left(\frac{\mathcal{I}(m_q, n_{q+1}) - \mathcal{I}(m_q, n_{q-1})}{\mathcal{I}(m_{q+1}, n_q) - \mathcal{I}(m_{q-1}, n_q)}\right) \tag{8}$$

For an image, the final PHOG descriptor is obtained by concatenating all HOG vector at each pyramid level for each p_{th} particle as $E_h^p = H_1, H_2...H_{\hat{N}_b}$. This depicts the spatial information for the image. Each HOG is normalized to obtain the final vector considering all the pyramid level. Zero level represents original image corresponding to b bins of each b HOG vector, level 1 by a $4b$ vector and the n_{th} level is represented as $\sum_{d=0}^{n} 2^{2d} = \sum_{d=1}^{n} 4^{2d}$. For each particle, Edge histogram (F_g) is determined by normalizing the value of the so obtained PHOG descriptor.

Similarity between each cue histogram (F_l) for p_{th} particle and the corresponding reference histogram dictionary $\hat{D}_r$ is determined using Bhattacharya's distance. For this, Bhattacharya's coefficient is calculated using (9).

$$\beta_l^p(\hat{D}_{r,l}, F_l^p) = \sum_{n=1}^{\hat{N}} \sqrt{(\hat{D}_{r,l})^n \times (F_l^p)^n}, \quad l \in i, e, g \tag{9}$$

This Bhattacharya's coefficient is used to calculate the Bhattacharya's distance [5] between cue histogram (F_l) for p_{th} particle and the corresponding reference histogram $\hat{D}_r$ in the dictionary using (10).

$$D_l^p(\hat{D}_{r,l}, F_l^p) = \sqrt{1 - \beta_l^p(\hat{D}_{r,l}, F_l^p)}, \qquad l \in i, e, g \tag{10}$$

Bhattacharya's distance corresponding to each component of dictionary is averaged. Further, the likelihood for each cue for p_{th} particle is calculated using (11).

$$\gamma_l^p(\hat{D}_{r,l}, F_l^p) = \frac{1}{\sigma_l\sqrt{2\pi}} e^{-\frac{D_l^p(\hat{D}_{r,l}, F_l^p)^2}{2\sigma_l^2}}, \quad l \in i, e, g \tag{11}$$

where, σ_l represents the Standard deviation for Gaussian noise for each cue. The next section will detail the adaptive multi-cue fusion model.

3.1.3 Adaptive multi-cue fusion

Each cue likelihood is subjected to adaptive fusion model to obtain the fused weight $\hat{W}_f$. Non-linear ranking based score fusion method has been proposed for adaptive fusion of the cues. This model boosts the important particles and suppresses the unimportant particles automatically. For this, the likelihoods obtained for each cue through (11) for p_{th} particle are assigned as initial weights using (12).

$$\hat{C}_{l,p} = \gamma_l^p(\hat{D}_{r,l}, F_l^p), \quad p = 1, 2, ...N \tag{12}$$

If N to be total number of particles and $l \in i, e, g$. Then, the individual weights for each particles of cue can be represented using (13).

$$\hat{C}_{l,N} = \begin{bmatrix} \hat{C}_{i1} & \hat{C}_{i2} & \dots & \hat{C}_{iN} \\ \hat{C}_{e1} & \hat{C}_{e2} & \dots & \hat{C}_{eN} \\ \hat{C}_{g1} & \hat{C}_{g2} & \dots & \hat{C}_{gN} \end{bmatrix} \tag{13}$$

Context reliability value for each cue is calculated at $t-1$ as $\hat{r}_t^l = (\hat{r}_{t-1}^i, \hat{r}_{t-1}^e, \hat{r}_{t-1}^g)$. These values are multiplied with each row corresponding to each cue to obtain (14).

$$CS_{l,N}^{\hat{f}} = \begin{bmatrix} \hat{r}_{t-1}^i * \hat{C}_{i1} & \hat{r}_{t-1}^i * \hat{C}_{i2} & \dots & \hat{r}_{t-1}^i * \hat{C}_{iN} \\ \hat{r}_{t-1}^e * \hat{C}_{e1} & \hat{r}_{t-1}^e * \hat{C}_{e2} & \dots & \hat{r}_{t-1}^e * \hat{C}_{eN} \\ \hat{r}_{t-1}^g * \hat{C}_{g1} & \hat{r}_{t-1}^g * \hat{C}_{g2} & \dots & \hat{r}_{t-1}^g * \hat{C}_{gN} \end{bmatrix} \tag{14}$$

After this, particles are ranked based on the obtained score $\hat{C}^{re}_{CN}$ for each cue individually. For each particle, the rank is obtained as $R_j = R_N, R_{N-1}...R_1$. In each row, particle with highest score is ranked as the highest rank (R_N) and the particle with lowest score is given lowest rank (R_1). The rank matrix $W_{l,N}$ is obtained using (15).

$$W_{l,N} = \begin{bmatrix} R_j * \hat{C}^{\hat{r}}_{i1} & R_j * \hat{C}^{\hat{r}}_{i2} & \dots & R_j * \hat{C}^{\hat{r}}_{iN} \\ R_j * \hat{C}^{\hat{r}}_{e1} & R_{ij} * \hat{C}^{\hat{r}}_{e2} & \dots & R_j * \hat{C}^{\hat{r}}_{eN} \\ R_j * \hat{C}^{\hat{r}}_{g1} & R_j * \hat{C}^{\hat{r}}_{g2} & \dots & R_j * \hat{C}^{\hat{r}}_{gN} \end{bmatrix} \tag{15}$$

These scores are normalized row-wise using the min-max normalization. The normalized score matrix is represented by (16.)

$$\bar{W_{l,N}} = \begin{bmatrix} W_{i1} & W_{i2} & \dots & W_{iN} \\ W_{e1} & W_{e2} & \dots & W_{eN} \\ W_{g1} & W_{g2} & \dots & W_{gN} \end{bmatrix} \tag{16}$$

The normalized score are subjected to $f : \bar{W_{l,N}} \rightarrow W^p_{fus}$ to obtain the fused score. $W^p_{fus} = [f(\bar{W_{i1}}, \bar{W_{e1}}, \bar{W_{g1}}), ...f(\bar{W_{iN}}, \bar{W_{eN}}, \bar{W_{gN}})]$. The function f for p_{th} particle is given by (17). Hence, using this, each particle is assigned weight after its evaluation over three cues.

$$\begin{aligned} W^p_{fus} = f(\bar{W_{ip}}, \bar{W_{ep}}, \bar{W_{gp}}) &= \frac{\bar{W_{ip}}}{1 + \bar{W_{ip}} \times \bar{W_{ep}} \times \bar{W_{gp}}} \\ &+ \frac{\bar{W_{ep}}}{1 + \bar{W_{ip}} \times \bar{W_{ep}} \times \bar{W_{gp}}} + \frac{\bar{W_{gp}}}{1 + \bar{W_{ip}} \times \bar{W_{ep}} \times \bar{W_{gp}}} \end{aligned} \tag{17}$$

Next, the fused weight are passed through a non-linear function to boost the concordant cues and suppress the discordant cues using (18).

$$\hat{W}^p_{f,t} = \frac{e^{W^p_{fus}} - e^{-W^p_{fus}}}{e^{W^p_{fus}} + e^{-W^p_{fus}}} \tag{18}$$

Where, $\hat{W}^p_{f,t}$ is the final weight assigned to p_{th} particle at time t. These final weights are further subjected to an outlier detection mechanism to classify the particles into important particles ($\hat{I}_y$) and unimportant particles ($\hat{U}_x$). Unimportant particles which are mainly affected due to environmental variations are termed as outliers. These particles are detected by the outlier detection mechanism using (19).

$$\hat{U}_x = \quad \hat{W}^p_f, \qquad where \quad \hat{W}^p_f < \tau_s \tag{19}$$

$$\hat{I}_y = \quad \hat{W}^p_f, \qquad where \quad \hat{W}^p_f \geq \tau_s \tag{20}$$

where, τ_s is a defined threshold , $|x \cup y = N|$, $\hat{U}_x \in [X_{t,x}, Y_{tx,}]$ and $\hat{I}_y \in [X_{t,y}, Y_{t,y}]$. These outliers are further subjected to Butterfly optimization algorithm (BOA) [2] based resampling technique, which has been discussed in the next section.

3.2 Optimum resampling approach

BOA [2] is a nature-inspired meta-heuristic approach which disperses the outliers in the high likelihood area by its two parameters namely sensor modality ($\hat{m}_o$) and switch

probability ($prob$). Sensor modality controls the search space and determines the convergence speed. Switch probability used to switch between the local search and global search. BOA initializes the outliers as butterfly population (x) and their global position in the solution space is updated using (21).

$$U_x = \begin{cases} \hat{U}_x + (rand^2 \times b' - \hat{U}_x) \times \mu_x & if \quad rand \leq prob \\ \hat{U}_x + (rand^2 \times \hat{U}_j - \hat{U}_k) \times \mu_x, & otherwise \end{cases} \tag{21}$$

Here, $j, k \in x$, $rand \in [0, 1]$ and $U_x \in [X_{x,t}, Y_{x,t}]$. b' represents the best global position of the particles in the iteration. μ_x is calculated as: $\mu_x = \hat{m}_o I^a$. Here, I is correlated with the current position of the particles, $\hat{m}_o$ and a are the search space controlling parameters and varies from 0 to 1.

Final centroid G_t of the target is estimated using fused weight $\hat{W}_f^p$ to that of outliers weight $\hat{W}_x$ and the important particles weight $\hat{W}_y$ using (22).

$$G_t = \frac{\sum_x \hat{W}_x \times U_x + \sum_y \hat{W}_y \times \hat{I}_y}{\sum_{p=1}^{N} W_p}, \quad |x \cup y = N| \tag{22}$$

where, $G_t = [X_t, Y_t]$ and $U_x, \hat{I}_y$ are the state determined by outliers and important particles respectively. Further, in order to ensure the quick adaptation of the proposed method during dynamic environment the reliability values are determined for each cue. Each cue reliability is estimated by calculating the L2-norm distance between the final estimated state and the state estimated through individual cue. L2-norm distance is calculated using (23).

$$d_{t,l} = ||G_t - G_t^l|| = \sqrt{(X_t - X_t^l)^2 + (Y_t - Y_t^l)^2} \quad \forall \quad l \in i, e, g \tag{23}$$

where, (X_t, Y_t) is the centroid of the final estimated state (G_t) and (X_t^l, Y_t^l) is the centroid of state determined through individual cue at time t. Using $d_{t,l}$ the cue reliability is calculated by (24).

$$\hat{r}_t^l = \frac{\tanh(-u(d_{t,l}) + h)}{2} + 0.5 \quad \forall \quad l \in i, e, g \tag{24}$$

where, u and h are constants. The context cue reliability $(\hat{r}_t)$ is calculated at state t and used for discounting the particles at state $t + 1$. These reliability values leads to adaptive fusion of color, texture and edge likelihoods in the update model of the proposed tracker. Important particles $(\hat{I}_y)$ and Outliers particles $(\hat{U}_x)$ are updated at each state. Also, reference dictionary is updated by selective replacement of the important particles. This process is repeated iteratively to keep target's track until it is visible in the scene. Once it is lost, the whole tracker is re-initialized to re-detect the target. The pseudocode of the proposed tracker is illustrated as Algorithm 1. Details about the experimental validation of the proposed tracker are discussed in the next section.

Algorithm 1 Proposed tracker model.

Function `Tracker()`:
  for *At time step t* **do**
    if $t == 1$ **then**
      Tracker intialized with N particles
      Calculate reference dictionary $\hat{D}_{r,i}$, $\hat{D}_{r,e}$, $\hat{D}_{r,g}$
      $\hat{r}_i, \hat{r}_e, \hat{r}_g \leftarrow 1$
    else
      Evolve p particle using (1)
      for *j from 1 to n* **do**
        Determine $\gamma_i, \gamma_e, \gamma_g$ using (11)
        Calculate $\hat{W}_f^p$ using (18)
        Determine $\hat{U}_x$ and $\hat{I}_y$ using (19), (20)
        for *c from 1 to x* **do**
          $rand \leftarrow U[0, 1]$
          calculate $\mu_x \leftarrow \hat{m}_o I_a$
          $\hat{m}_o, a \leftarrow U[0, 1]$
          if $rand \leq prob$ **then**
            $U_x = \hat{U}_x + (rand^2 \times b^{'} - \hat{U}_x) \times \mu_n$
          else
            $U_x = \hat{U}_x + (rand^2 \times + \hat{U}_x - \hat{U}_x) \times \mu_n$
          end
        end
        Update U_x
      end
    end
    Obtain updated G_t using (22)
    Update $\hat{r}^i, \hat{r}^e, \hat{r}^g$ using (24)
  end
  return $G_1, G_2, ..., G_t$

4 Experimental validation and discussion

In order to analyze the performance of the proposed approach, we have chosen publicly available sample video sequences taken from OTB-100 dataset [37] and VOT dataset [18]. The chosen video sequences are rich in various environmental challenges such as background clutters, scale variations, deformation, full or partial occlusion, illumination variations, fast motion and motion blur. Results of the proposed tracker are compared against 13 other state-of-the-art viz. ASLA [17], MTT [43], CT [41], FRAG [1], IVT [26], MIL [3], WMIL [40], DFT [28], PF-PSO [42], PF [14], CSPF [34], MCPF [44] and STAPLE [4]. The proposed tracker is implemented in Python2 on a $i5$ quadcore 2.4 GHz processor with 8 GB RAM. Initially, the target is detected through GMM subtraction model [29] and N= 49 particles are instantiated on it. The dictionary is updated by selective replacement of the important particles. Tracker is executed 5 times iteratively to handle the probabilistic nature of the particle filter. Results depicted are average obtained over 5

iteration of the algorithm. The next section detail about the evaluation metrics for the proposed tracker.

4.1 Evaluation metrics

To prove the effectiveness and accuracy of the proposed tracker over other state-of-the-art, robust evaluations metrics viz. Center location error, F-Measure [19], AUC (Area Under Curve), DP (Distance Precision), OP (Overlap precision), Success plots and Precision plots [37] are considered. Center Location Error (CLE) is the Euclidean distance between the tracker's bounding box (BB) and the ground truth BB. F-Measure is calculated as the harmonic mean between the precision (p_{rc}) and recall (RC). Precision (p_{rc}) is the ratio of the overlap of the tracker's BB and the ground truth BB w.r.t. tracker's BB. Recall (RC) is the ratio of the overlap of the tracker's BB and the ground truth BB w.r.t. ground truth BB. Success plot depicts the success score at different overlap thresholds and mean precision at multiple center location error is depicted by the precision plots. AUC is average of success rate determined at different overlap threshold and DP is average precision score calculated at different location thresholds. Overlap precision score is the percentage of frames where BB is above a threshold.

4.2 Attribute based evaluation

In this section, the performance of the proposed tracker has been evaluated on 11 attributed challenges namely, illumination variations,background clutters, fast motion & motion blur, scale variations & deformation, in-plane & out-of-plane rotations and full or partial occlusion. The details of the tracking challenges and the considered video sequence has been tabulated in Table 1. Sample tracking frames from the considered challenging video sequences under various attributed challenges are depicted in Fig. 2.

Table 1 Tracking challenge and considered video sequence

Tracking challenge	Video sequences
Illumination variations	Fish, Human8, Skating1, Singer1, Car2, Soccer1, Singer2, Tiger, Shaking, Basketball, Crossing
Background clutter	Basketball, Car2, Pedestrian1, Football, Shaking, Singer2, Bolt2, MountainBike, Skating1
Fast Motion and Motion Blur	CarScale, Pedestrian1, Jumping, Soccer1, Jogging1, Tiger, Human7, Crossing
Scale variations & Deformation	CarScale, Shaking, Dancer, Human8, Walking, Jogging1, Human7, Bolt2, Skating1, Basketball, Pedestrian1
In-plane & Out-of-plane rotation	Basketball, CarScale, Pedestrian1, Football, Shaking, Singer2, Skating1, Soccer1, Tiger, Dancer, MountainBike
Full or partial occlusion	Football, CarScale, Subway, Walking2, Jogging1, Singer1, Jogging2, Tiger, Basketball

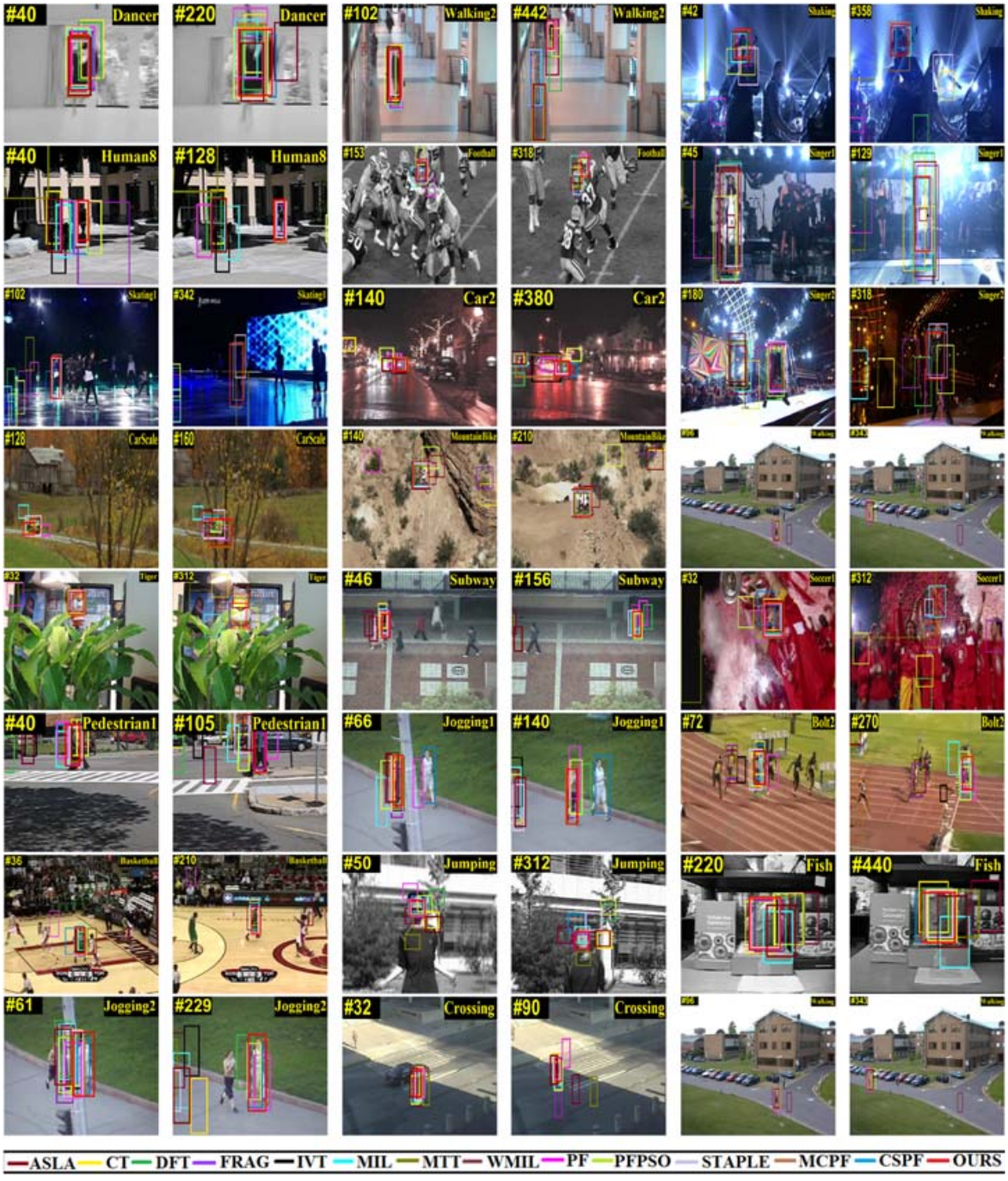


Fig. 2 Representative frames from the considered challenging video sequences. Each frame has been labeled with # frame number and video sequence name on the left and right corner, respectively

Illumination Variations: In *Car2* sequence, at frame #380 most of the trackers have shown drift whilst Ours, Frag and STAPLE are able to manage the change in illumination in the whole sequence. At frame #128 of *Human8*, when there is a sudden variation in illumination, most of the trackers lost the target but the proposed tracker with STAPLE and MCPF are able to locate the target till the end. Here, the PHOG cue enhances proposed tracker's performance by compensating for color and LBP cues. Another challenging sequence under this challenge is *Shaking*. Initially, all the trackers are able to keep track of the target with the marginal error. However, at frame #358 CT, FRAG, PF-PSO, IVT and MTT has lost the target. Ours with MIL, WMIL, ASLA, and MCPF are able to keep track the target. This is due

to the automatic boosting of good particles and suppression of the unimportant particles by the proposed fusion model. In *Singer2*, at frame #180, only our tracker, CSPF and STAPLE are able to locate the target while others have shown deflection. In *Singer1* Ours, CSPF and MCPF have shown substantially good performance in comparison to other trackers. *Fish* and *Soccer1* sequences have abrupt illumination variations in which Ours and MCPF have much better results in comparison to other trackers. In sum, the proposed tracker performs with substantially low error under this challenge which is due to the considered complementary cues in the appearance model. In addition, consistent updating of the reference dictionary with the important particles improves the performance of the proposed tracker under illumination variations. Also, the proposed tracker has achieved highest AUC of 0.696 depicted in Fig. 3a and average DP score of 0.761 depicted in Fig. 4a under the challenge.

Background Clutters: The sequences viz. *Crossing*, *Car2*, *Shaking* and *Singer2* has similar background to that of the target . For *crossing* sequence, when there is similar background at frame #90 ASLA, FRAG, MTT and PF lost the target whilst Ours, CSPF, MCPF and STAPLE locate the target with very small error. *Pedestrian* sequence has background clutter at frame #105 when the target suffers from background clutter then CT, FRAG, PF-PSO, MCPF, and MIL have lost the target and start tracking the similar objects in the background but Ours track the target successfully. This is due to the boosting of important particles and suppression of unimportant particles by adaptive multi-cue fusion. In *Tiger* sequence, trackers viz. CT, ASLA, FRAG, MTT and MCPF could not able to handle the tracking challenge. *Walking* sequence, MTT, STAPLE and MCPF have shown better performance in comparison to the proposed tracker. This may be due to the false update of the particles due to similar background during tracking. *Bolt2* and *Skating1* sequences have several pose variations with similar background. Ours keep track of the target in the whole sequence. This owes by the boosting of robust complementary LBP and PHOG features over color. In sum, the robustness of the proposed tracker under background clutter is mainly due to the proposed adaptive fusion model which boost the important particles and suppress the unimportant particle during the challenge. Unimportant particles are detected as outliers by the proposed outlier detection mechanism and their position is further improved by the proposed resampling method. In addition, cue reliability is also discounted for the quick adaptation of target during background clutters. Under Background clutters challenge, the proposed tracker has attained highest AUC of 0.662 illustrated in Fig. 3b and average DP score of 0.668 illustrated in Fig. 4b.

Fast Motion and Motion Blur: Generally, fast motion and motion blur occur in a sequence either by abrupt target's motion or camera movement. *Pedestrian1* and *Human7* sequences has motion blur due to sudden camera movement. Ours has handled the challenge gracefully in these sequences. It is due the rotational factor in the state model and the detection of the affected particles by the proposed outlier detection mechanism. In *Tiger* and *Soccer1* sequences, target is moving very fast and trackers viz. STAPLE, CSPF, MIL, MTT and ASLA are not able to keep track of the target. But, Ours and MCPF has shown considerably good performance. *Jumping* sequence has tough tracking scenario with both fast motion and motion blur. The proposed tracker performs well in comparison to other trackers by the proposed fusion model which suppresses the affected particles and boosts the good particles. Under fast motion & motion blur challenge, the proposed tracker has achieved highest AUC of 0.664 as shown in Fig. 3c and highest average DP score of 0.753 as shown in Fig. 4c.

Scale Variations and Deformation: For sequence *Dancer*, the proposed tracker with MTT, IVT, CSPF, STAPLE and MCPF are able to handle the target deformation due to scale variations. However, MIL and WMIL are slightly deflected from the target. At frame #140, in *MountainBike* sequence target has scale variations and deformation due to its sudden

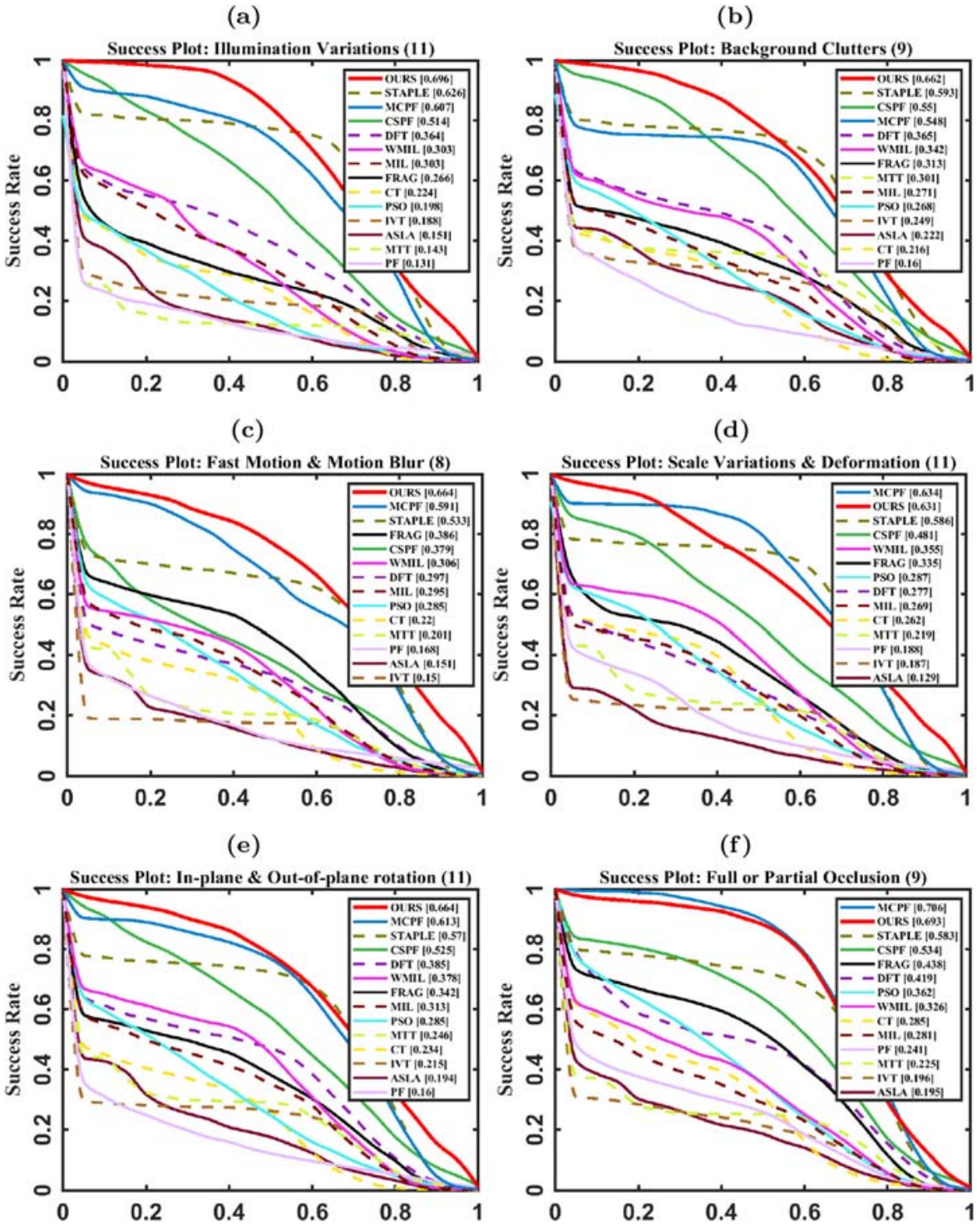


Fig. 3 Success plot under visual tracking challenges as **a** Illumination variations **b** Background Clutters **c** Fast Motion and Motion Blur **d** Scale variations and Deformation **e** In-plane and Out-of-plane rotation **f** Full or partial occlusion. Legend includes Area Under Curve (AUC) in the bracket

movement which PF, FRAG, CT and WMIL could not able to handle. Our, ASLA, MIL, MCPF and STAPLE has handled this variation with minimal error. In *CarScale* sequence, at frame #160 most of the tracker lost the target as these trackers appearance model is inefficient to handle large scale variation. But, MCPF and STAPLE have performed marginally better than the proposed tracker. In *Basketball*, ASLA, MIL, CT, MTT and DFT lost the target whilst Ours with FRAG, WMIL,CSPF and MCPF have shown better performance.

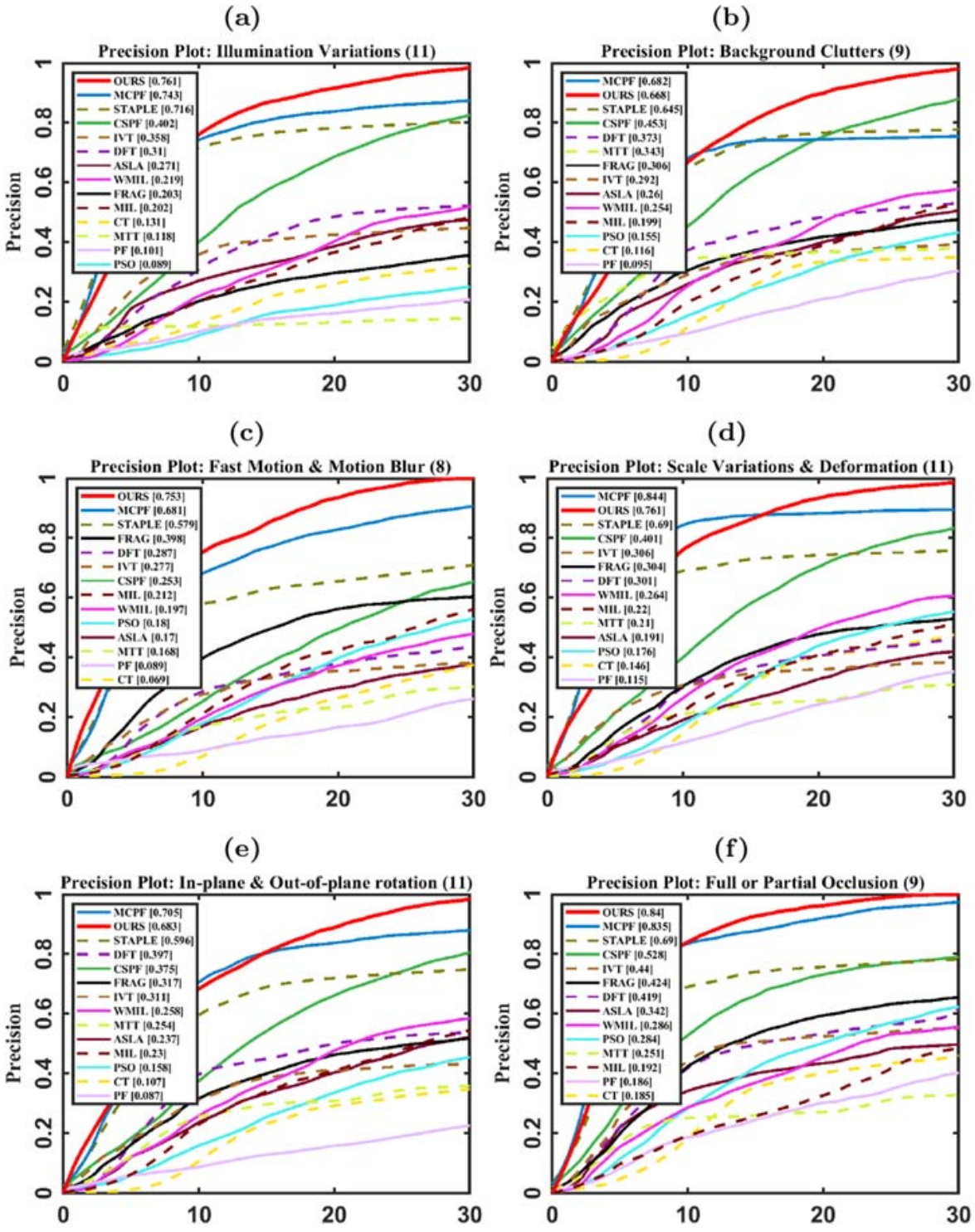


Fig. 4 Precision plot under visual tracking challenges as **a** Illumination variations **b** Background Clutters **c** Fast Motion and Motion Blur **d** Scale variations and Deformation **e** In-plane and Out-of-plane rotation **f** Full or partial occlusion. Legend includes average Distance Precision (DP) score in the bracket

This is due the scale-invariant LBP feature in the proposed tracker which handled the target variations. Ours has shown better performance in comparison to other trackers in *Walking2* throughout the whole sequence. It may be due to the scaling component of the state model which has catered the scale variation in the sequence. In sum, the proposed tracker is robust under scale variations which is mainly attributed by the incorporation of scale-invariant LBP feature in the appearance model. Moreover, the incorporation of the scaling factor in the

random walk model of the state model has strengthened our method under this challenge. Under the challenge, the proposed tracker has attained the second highest AUC score of 0.631 depicted in Fig. 3d and second highest average DP score of 0.761 depicted in Fig. 4d.

In-plane and Out-of-plane rotation: These rotations appear frequently in the sequence when the target moves in or out of the image plane. In sequences, *Basketball*, *Pedestrian1* and *Football* the target rotates in and out of the image plane suddenly. The proposed tracker handled these variations by adaptive boosting of the LBP and PHOG features over color. In *Singer2* and *Soccer1*, most of the trackers are not able to keep track of the target. However, Ours has shown substantially better performance. Ours, CSPF, MCPF and STAPLE has shown good tracking results in comparison to other trackers in the sequence *Tiger* and *Shaking*. The proposed tracker is able to handle the challenge to a great extent. The rotational component in the state model of the proposed tracker and the complementary cues, color, LBP and HOG in the appearance model of the tracker enhances its performance under the challenge. Moreover, the proposed tracker outperforms the other trackers by achieving the highest AUC of 0.664 illustrated in Fig. 3e and second highest DP score of 0.683 as illustrated in Fig. 4e under the challenge.

Full or Partial Occlusion: Severe occlusion has been noticed in the *Jogging1* and *Jogging2* sequence, when the pillar occluded the moving target. Most of the trackers drift away from the target. The proposed tracker and MCPF are able to track the target after it recovers from the full occlusion. *Soccer1* sequence has multiple full and partial occlusion challenge. At frame #312 when the target is partially occluded by the similar objects most of the tracker are deflected from the target. However, the proposed method gracefully handle the challenge and track the target successfully. This is due the complementary cues exploited in the appearance model. There are multiple full and partial occlusion of the target in *Subway* sequence. At frame #156 when the target is recovered from occlusion, then ASLA, WMIL and MTT totally lost the target. Only Ours, CSPF, MCPF and STAPLE can keep track of the target even after occlusion. Generative based trackers are not able to handle occlusion as their appearance model is inefficient in handling this challenge. For *Football* sequence, when target is occluded by the other players then Ours with MCPF track the target successfully. This is due to the selective boosting of the color and LBP features and suppression of the PHOG by the proposed fusion model during occlusion. In *Tiger*, the proposed tracker with CSPF has shown better tracking results in comparison to other trackers. The adaptiveness of the proposed tracker under full or partial occlusion primarily due to the proposed fusion model which suppress the unimportant particles whose weight decreases during occlusion and boost the other important particle. These unimportant particles are detected by the outlier mechanism and hence, prevent erroneous update of the tracker during occlusion. These particles are further subjected to the proposed resampling method in order to enhance their contribution to the state estimation. Also, as shown in Figs. 3f and 4f the proposed tracker has achieved AUC of 0.693 and average DP score of 0.840, respectively under full or partial occlusion challenge.

4.3 Analysis of feature contribution

The performance of the proposed method considering individual feature and all possible multiple combination of the features in terms of CLE has been tabulated in Table 2. Challenging video sequences are considered in order to prove the robustness of exploiting three complementary multi-cue in the appearance model of the proposed tracker. Results infer that single cue is inefficient in handling the various environmental variations. Multiple cues namely, color and LBP, LBP and PHOG or PHOG and color when integrated together then,

Table 2 Comparison of different contribution of features

Video sequence	Color	LBP	PHOG	Color + LBP	LBP + PHOG	Color + PHOG	Ours
Car2	41.82	25.06	18.29	28.5	19.61	41.12	5.15
MountainBike	49.36	165.76	85.64	18.36	16.22	15.81	7.69
Jogging1	29.12	26.77	24.55	11.24	13.13	13.55	6.85
Crossing	11.77	11.85	75.51	13.84	13.66	34.13	4.77
Shaking	121.48	183.42	49.08	29.6	31.97	27.51	18.25
CarScale	39.14	94.88	36.83	18.44	26.93	17.44	10.85
Soccer1	199.9	118.2	174.49	25.09	25.87	28.04	11.31
Pedestrian1	43.16	43.91	25.65	22.09	20.74	17.96	12.26
Human8	29.43	68.85	29.02	29.43	18.71	18.84	9.67
Dancer	34.23	69.06	40.43	17.8	16.74	26.16	4.1
Subway	46.98	44.98	38.56	10.5	11.82	7.98	6.19
Tiger	219.4	107.55	210.91	37.69	39.69	43.7	7.28

performance of the tracker improves marginally. However, when these cues are combined using the proposed fusion approach, performance of the tracker substantially enhanced by addressing the various environmental challenges.

4.4 Computational complexity analysis

In order to analyze the computational complexity of the proposed tracker in real-time, we have calculated the average computational speed score of the proposed tracker and compared the results with other trackers. Average speed score is calculated by average fps obtained over all the video sequences. Table 3 tabulates the computational speed analysis of the proposed method in comparison to the other methods. It has been evident from the table that the proposed tracker has shown real-time performance with high accuracy.

The proposed tracker has shown comparable performance against other methods and outperforms trackers namely, ASLA, MIL, DFT, WMIL, CT, FRAG, STAPLE, MCPF, PSO and CSPF in terms of accuracy by 89.41%, 45.65%, 27.64%, 35.56%, 81.73%, 83.29%, 16.88%, 13.12%, 79.57% and 48.77% respectively. Although, MCPF has shown comparable performance to the proposed tracker but it is computationally slow.

4.5 Overall performance evaluation

In order to evaluate the overall performance of the proposed approach, Tables 4 and 5 have tabulated the average CLE (in pixels) and average F-Measure of the proposed tracker along with 13 other state-of-the-art, respectively. The proposed tracker has achieved an average mean CLE of 6.89 (in pixels) and average mean F-Measure of 0.786 when evaluated on

Table 3 Tracking speed analysis (FPS)

Tracker	ASLA	MIL	DFT	WMIL	CT	FRAG	STAPLE	MCPF	PSO	CSPF	OURS
Speed (FPS)	16	7	9	18	19	14	18.4	0.57	15	21	20
Accuracy	0.0745	0.3823	0.5090	0.4533	0.1285	0.1175	0.5847	0.6112	0.1437	0.3604	0.7035

Table 4 Comparison results for average Centre Location Error (CLE). First, Second and Third results are highlighted

Challenge	ASLA	MTT	MIL	DFT	WMIL	IVT	CT	FRAG	PSO	PF	MCPF	STAPLE	CSPF	OURS
Illumination variations	70.93	162.236	62.16	63.66	63.32	79.07	83.00	78.75	69.43	178.317	23.36	29.27	17.53	7.87
Scale variations & deformation	87.11	108.78	56.52	66.69	64.99	67.79	73.74	60.26	48.06	108.74	26.27	34.44	17.74	7.68
Fast motion & motion blur	69.20	96.25	42.93	55.03	53.28	74.12	63.11	49.85	52.30	98.97	10.82	30.71	27.19	7.03
Background clutters	73.06	111.88	64.17	68.28	81.35	78.98	103.13	87.98	56.88	135.82	52.24	33.11	14.56	9.22
Full or partial occlusion	77.13	142.25	61.92	33.81	64.03	66.66	59.95	38.28	45.36	131.111	5.92	33.28	19.04	6.31
In-plane and out-of-plane rotation	74.47	124.58	60.50	67.15	70.11	81.64	96.49	83.90	69.18	149.31	24.24	34.60	18.79	8.88
Overall	76.92	108.08	50.49	47.91	66.43	65.59	70.35	61.10	46.75	110.27	**24.19**	28.26	**16.55**	**6.89**

Table 5 Comparison results for average F-Measure. First, Second and Third results are highlighted

Challenge	ASLA	MTT	MIL	DFT	WMIL	IVT	CT	FRAG	PSO	PF	MCPF	STAPLE	CSPF	OURS
Illumination variations	0.193	0.153	0.378	0.434	0.389	0.205	0.278	0.319	0.259	0.148	0.708	0.701	0.634	0.804
Scale variations & deformation	0.155	0.259	0.332	0.339	0.437	0.200	0.332	0.411	0.371	0.235	0.739	0.661	0.587	0.734
Fast motion & motion blur	0.186	0.242	0.367	0.350	0.376	0.156	0.277	0.466	0.365	0.203	0.700	0.602	0.464	0.764
Background clutters	0.271	0.334	0.332	0.439	0.417	0.277	0.268	0.367	0.344	0.294	0.630	0.671	0.669	0.772
Full or partial occlusion	0.241	0.254	0.347	0.503	0.401	0.222	0.365	0.526	0.463	0.197	0.815	0.662	0.632	0.799
In-plane and out-of-plane rotation	0.244	0.289	0.384	0.457	0.462	0.235	0.293	0.409	0.366	0.193	0.715	0.647	0.639	0.769
Overall	0.232	0.294	0.402	0.450	0.385	0.288	0.338	0.425	0.381	0.258	**0.716**	**0.681**	0.628	**0.786**

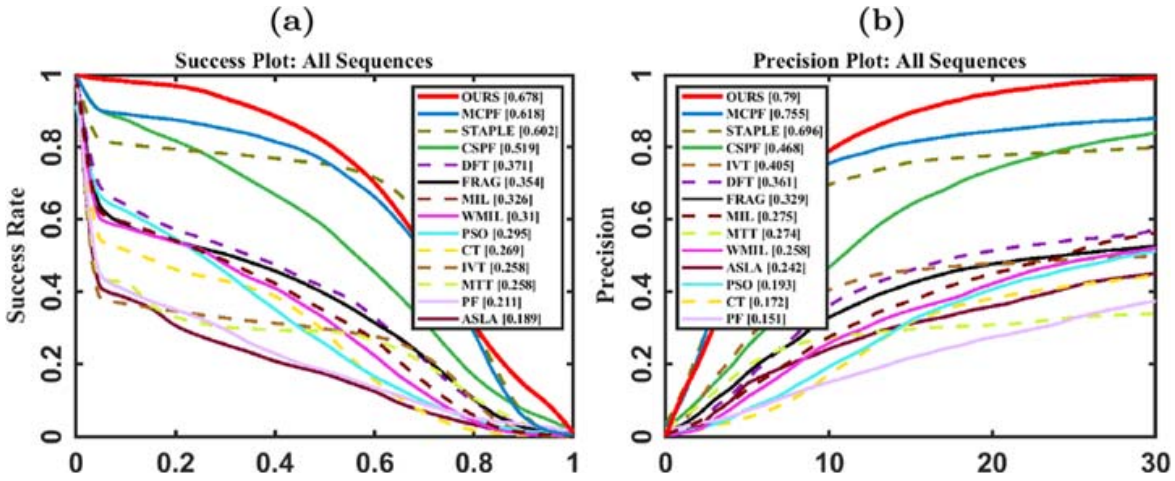


Fig. 5 Overall performance evaluation of the proposed tracker with other state-of-the-art: **a** Success plot **b** Precision plot. Legend includes the AUC for success plot and average DP for precision plot respectively, in the brackets

challenging video sequences. Overall success and precision plots are depicted in Fig. 5. Success plot illustrates that the proposed tracker has achieved high AUC score of 0.678 and precision plot depicts the high precision score of 0.790 at 10px for our tracker.

In sum, low CLE value and high F-Measure of the proposed tracker reveal its robustness during various environmental challenges. Also, Success vs overlap threshold plots and precision vs CLE threshold (in pixel) plots indicate that the proposed tracker is able to locate the target precisely. The drawbacks of particle filter viz. particle degeneracy and sample impoverishment are handled by the exploited butterfly optimization based resampling technique. Our tracker is computationally efficient as it resamples only the outlier particles. Unimportant particles which contribute less in the state estimation or affected due to illumination variation, occlusion, background clutters, fast motion and in-plane and out-of-plane rotations are detected as outliers by the outlier detection mechanism. In addition, the proposed adaptive fusion model integrates the multi-cue weight through context cue reliability which ensure consistent updation of the tracker. The fusion model boosts the important particles and suppresses the unimportant particles for the final state estimation. Our tracker can deal with various tracking challenges to a great extent. On the other hand, generative trackers viz. IVT, FRAG and MTT were tend to drift under full occlusion and background clutters. Discriminative trackers viz. MIL and CT were prone to illumination and scale variations. Particle filter based trackers viz. PF and PF-PSO had limited performance due to the limitation of the resampling method. WMIL had exploited random Haar features and not able to handle the object's deformation and rotation. STAPLE was not able to cater dynamic appearance variations. CSPF was susceptible to fast motion and motion blur. MCPF had shown comparable performance but it is not suitable for real-time tracking. However, Our tracker achieves real-time performance with high accuracy under various environmental variations.

5 Conclusion

In this paper, we have proposed a multi-cue object tracking solution under particle filter framework by adaptive fusion of the cues. Adaptive fusion model boosts the important particles and suppresses the unimportant particles to handle occlusion, background clutters and motion blur. An outlier detection mechanism is also proposed to disparate the

low performing particles affected due to environmental variations. These particles are relocated to high likelihood area by the proposed resampling method. This resampling methods not only address the drawbacks of particle filter but also enhances tracker's performance. Also, a robust multi-cue appearance model is proposed to handle the illumination variations and scale variations. Incorporation of scaling and rotation factor in state vector of the state model caters the change in scale, fast motion, in-plane and out-of-plane rotation of the target. Context sensitive cue reliability discounted the particles for the quick adaptation of the proposed approach to the dynamic environment. In addition, the particle degeneracy and sample impoverishment are handled by the butterfly optimization based resampling technique. Qualitative and quantitative validation evaluates the superiority of the proposed method against 13 other state-of-the-art trackers.

In future, we would like to introduce more cues at each stage for precise localization of the target. Object's pose variations, deformation with scale variations are other directions which can be explored further. Multimodal information with source importance parameters can also be investigated for object tracking.

Compliance with Ethical Standards

Conflict of interests All authors declare that they have no conflicts of interest.

References

1. Adam A, Rivlin E, Shimshoni I (2006) Robust fragments-based tracking using the integral histogram. In: 2006 IEEE CS conf. on comp visi and pattern recog, vol 1, pp 798–805
2. Arora S, Singh S (2019) Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 23(3):715–734
3. Babenko B, Yang M-H, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632
4. Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr PHS (2016) Staple: complementary learners for real-time tracking. In: The IEEE conference on computer vision and pattern recognition (CVPR)
5. Bhattacharyya A (1943) On a measure of divergence between two statistical populations defined by their probability distributions. Bull Calcutta Math Soc 35:99–109
6. Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on image and video retrieval. ACM, pp 401–408
7. Brasnett P, Mihaylova L, Bull D, Canagarajah N (2007) Sequential Monte Carlo tracking by fusing multiple cues in video sequences. Image Vis Comput 25(8):1217–1227
8. Choe G, Wang T, Liu F, Hyon S, Ha J (2014) Particle filter with spline resampling and global transition model. IET Comp Vis 9(2):184–197
9. Dou J, Li J (2014) Robust visual tracking based on interactive multiple model particle filter by integrating multiple cues. Neurocomputing 135:118–129
10. Dou J-F, Li J-X (2014) Robust visual tracking base on adaptively multi-feature fusion and particle filter. Optik Int J Light Electron Optics 125(5):1680–1686
11. Fan Z, Ji H, Zhang Y (2015) Iterative particle filter for visual tracking. Signal Process Image Commun 36:140–153
12. Firouznia M, Faez K, Amindavar H, Koupaei JA (2018) Chaotic particle filter for visual object tracking. J Vis Commun Image Rep 53:1–12
13. Gao M-L, Li L-L, Sun X-M, Yin L-J, Li H-T, Luo D-S (2015) Firefly algorithm based particle filter method for visual tracking. Optik Int J Light Elect Optics 126:1705–1711
14. Gordon NJ, Salmond DJ, Smith AF (1993) Novel approach to nonlinear/non-gaussian bayesian state estimation. In: IEE proceedings F (radar and signal processing), vol 140. IET, pp 107–113
15. Han H, Ding Y-S, Hao K-R, Liang X (2011) An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking. Comp Maths App 62:2685–2695
16. Huan R, Bao S, Wang C, Pan Y (2018) Anti-occlusion particle filter object-tracking method based on feature fusion. IET Image Process 12(9):1529–1540

17. Jia X, Lu H, Yang M-H (2012) Visual tracking via adaptive structural local sparse appearance model. In: IEEE conf. on computer vision and pattern recognition, pp 1822–1829
18. Kristan M, Matas J, Leonardis A, Vojir T, Pflugfelder R, Fernandez G, Nebehay G, Porikli F, Čehovin L (2016) A novel performance evaluation methodology for single-target trackers. IEEE Trans Pattern Anal Mach Intell 38(11):2137–2155. https://doi.org/10.1109/TPAMI.2016.2516982
19. Lazarevic-McManus N, Renno J, Makris D, Jones GA (2008) An object-based comparative methodology for motion det. based on the f-measure. Comp Vis Image Un 111:74–85
20. Li T, Bolic M, Djuric PM (2015) Resampling methods for particle filtering: classification, implementation, and strategies. IEEE Signal Proc Mag 32(3):70–86
21. Li W, Wang P, Qiao H (2016) Top–down visual attention integrated particle filter for robust object tracking. Signal Process Image Commun 43:28–41
22. Li X, Hu W, Shen C, Zhang Z, Dick A, Hengel AVD (2013) A survey of appearance models in visual object tracking. ACM Trans Intell Sys Technol (TIST) 4(4):58
23. Lin SD, Lin J-J, Chuang C-Y (2015) Particle filter with occlusion handling for visual tracking. IET Image Process 9(11):959–968
24. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Machine Intell 24(7):971–987
25. Qian X, Han L, Wang Y, Ding M (2018) Deep learning assisted robust visual tracking with adaptive particle filtering. Signal Process Image Commun 60:183–192
26. Ross DA, Lim J, Lin R-S, Yang M-H (2008) Incremental learning for robust visual tracking. Int J Comput Vis 77:125–141
27. Sardari F, Moghaddam ME (2017) A hybrid occlusion free object tracking method using particle filter and modified galaxy based search meta-heuristic algorithm. App Soft Comp 50:280–299
28. Sevilla-Lara L, Learned-Miller E (2012) Distribution fields for tracking. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 1910–1917
29. Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: IEEE CS conf. on computer vision and pattern recognition, vol 2, pp 246–252
30. Walia GS, Ahuja H, Kumar A, Bansal N, Sharma K (2019) Unified graph-based multicue feature fusion for robust visual tracking. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2019.2920289
31. Walia GS, Kapoor R (2014) Intelligent video target tracking using an evolutionary particle filter based upon improved cuckoo search. Expert Syst Appl 41(14):6315–6326
32. Walia GS, Kapoor R (2016) Recent advances on multicue object tracking: a survey. Artificial Intell Rev 46:1–39
33. Walia GS, Kapoor R (2016) Robust object tracking based upon adaptive multi-cue integration for video surveillance. Multimed Tools Appl 75(23):15821–15847
34. Walia GS, Kumar A, Saxena A, Sharma K, Singh K (2019) Robust object tracking with crow search optimized multi-cue particle filter. Pattern Anal Applic: 1–17
35. Wang F, Lin B, Li X (2017) An ant particle filter for visual tracking. In: 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS). IEEE, pp 417–422
36. Wen Z, Peng Z, Deng X, Li S (2011) Particle filter object tracking based on multiple cues fusion. Procedia Eng 15:1461–1465
37. Wu Y, Lim J, Yang M-H (2013) Online object tracking: a benchmark. In: IEEE Conf on Comp Vis and Pattern Recog, pp 2411–2418
38. Xiao J, Stolkin R, Oussalah M, Leonardis A (2016) Continuously adaptive data fusion and model relearning for particle filter tracking with multiple features. IEEE Sensors J 16(8):2639–2649
39. Yin M, Zhang J, Sun H, Gu W (2011) Multi-cue-based camshift guided particle filter tracking. Expert Syst Appl 38(5):6313–6318
40. Zhang K, Song H (2013) Real-time visual tracking via online weighted multiple instance learning. Pattern Recogn 46:397–411
41. Zhang K, Zhang L, Yang M-H (2012) Real-time compressive tracking. In: European conf. on computer vision. Springer, pp 864–877
42. Zhang M, Xin M, Yang J (2013) Adaptive multi-cue based particle swarm optimization guided particle filter tracking in infrared videos. Neurocomputing 122:163–171
43. Zhang T, Ghanem B, Liu S, Ahuja N (2012) Robust visual tracking via multi-task sparse learning. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2042–2049
44. Zhang T, Xu C, Yang M-H (2017) Multi-task correlation particle filter for robust object tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4335–4343

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Ashish Kumar is currently pursuing his Ph.D. degree in computer science and engineering from Delhi Technological University (formerly DCE), New Delhi. He has completed his M.Tech degree in computer science and engineering from Guru Govind Singh Inderprastha University, New Delhi. He has published many research papers in various reputed national and international journals and conferences. His current research interests include object tracking, image processing and machine learning.



Dr. Gurjit Singh Walia received his Ph.D. degree in the field of computer vision from Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, and the M.E. degree in electronics from Punjab Engineering College, Chandigarh. He is working as Senior Scientist with Defence Research and Development Organization, New Delhi. His current research interests include machine learning, pattern recognition, and information security. He has published over 30 research papers in International Journals and Conferences. He is Fellow of IETE, Fellow of Institute of Engineering, Senior Member of IEEE, Member of CSI and CRSI. He is reviewer of various IEEE transactions, Elsevier & Springer journals.



Prof. Kapil Sharma is working as Professor and Head of Department of Information Technology, Delhi Technological University, Delhi. In 2011, he has completed Doctors Degree in Computer Science and Engineering from M.D. University, Rohtak (Haryana), India. He has obtained his B.E. and M.Tech. Degrees in Computer Science & Engineering from M. D. University, Rohtak and IASE, Rajasthan, India in 2000 and 2005 respectively. His research interests include system design, pattern recognition, computer vision and soft computing. He has published more than 100 research papers in referred journal and conference. He is reviewer for IEEE Transactions, Elsevier, Springer and Taylor & Francis journal.

Analog Integrated Circuits and Signal Processing (2020) 103:511–522
https://doi.org/10.1007/s10470-020-01657-2

MIXED SIGNAL LETTER

Resistorless current mode precision rectifier using EXCCII

Rupam Das[1] · Sajal K. Paul[2]

Received: 13 June 2019 / Revised: 5 April 2020 / Accepted: 13 April 2020 / Published online: 25 April 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
In this paper, a new current mode precision rectifier circuit is proposed using a single EXCCII and two nMOS transistors. The novelty of the circuit is that it can provide all the possible rectified outputs, namely ± full wave as well as ± half wave simultaneously without any alteration of topology. Moreover, it has low input impedance and high output impedance, which is suitable for fully cascadable operation. It is suitable for a very high frequency of operation. The operating frequency is found to be 125 MHz and above. The temperature, average DC voltage output, noise, power consumption, input dynamic range (± 500 μA), Monte-Carlo and total harmonic distortion analysis are carried out to check the performance quality. The conversion of sinusoidal signal to root-mean-square circuit is also included here as an application of the proposed rectifier The proposed circuit is simulated using Cadence ORCAD PSpice simulator with 0.18 μm CMOS technology parameters. Simulation results agree well with the theoretical analysis.

Keywords Rectifier · EXCCII based rectifier · Current mode rectifier · Precision rectifier · Sinusoidal to RMS

1 Introduction

A rectifier is a circuit which converts an alternating signal to a unidirectional signal. The conventional rectifiers are not suitable for applications where high precisions are required. High precision rectifiers are useful for a wide range of applications which include signal processing, conditioning, instrumentation, measurement and control of the low-level signals, AC voltmeters and ammeters, watt meters, linear function generators, sample and hold circuits, signal polarity detectors, radio frequency (RF) demodulators or amplitude modulated (AM) signal detectors, function fitting, error measurements, clipper circuits, frequency doubling, absolute value computation, and in-floor detectors used in ultrasonic.

Operational amplifier based rectifiers have restriction on the frequency range of operation due to limited slew rate and gain-bandwidth product. However, the implementation of precision rectifiers using current-mode approach has addressed those shortcomings to some extent. There are both voltage mode (VM) [1–10] and current mode (CM) [11–33] implementations using current mode approach. Since the proposed work is on current mode (CM) precision rectifier, a review on these rectifiers [11–33] is given to be precise. In [11] a novel CM full-wave rectifier (FWR) is presented. It uses one current differencing transconductance amplifier (CDTA), one grounded resistor and four diodes. In [12] another CDTA based FWR has been presented wherein one CDTA along with two diodes are used. Operating frequency for both [11, 12] is 5 MHz. The literature in [13] uses two analog building blocks (ABBs); one-second generation current conveyor (CCII) and one universal voltage conveyor (UVC) along with two diodes. Its operating frequency is low. Moreover, in [14] two topologies are presented. One of them uses two CCIIs and two diodes, whereas the other one uses one CCII, one UVC, and two diodes. They also require additional bias current and the operating frequency is 1 MHz. The implementations in [15, 19, 28–33] are not built around any standard ABB; however, uses nMOS transistors and [15] uses an additional two biases. The topology in [16] uses

✉ Sajal K. Paul
sajalkpaul@rediffmail.com

Rupam Das
das_rupam@rediffmail.com

[1] Electronics and Communication Engineering, Asansol Engineering College, Asansol, India

[2] Electronics Engineering, Indian Institute of Technology (ISM), Dhanbad, India

one CDTA, one resistor, and four diodes. Another CDTA based FWR has been presented in [17] using two CDTAs, one multiple output current follower (MCF), four diodes, and two grounded resistors. In [18] an MZC-CDTA along with two MOS transistors are used, but the operating frequency is 10 kHz only. The work reported in [20] uses only one CDTA, two MOS transistor as a switch and also requires three auxiliary bias voltages. In [21] two FWRs have been presented. First one is designed using one OTA, two grounded resistors and two diodes, whereas the second one uses one differential voltage current conveyor (DVCC), three grounded resistors, and two diodes. Both the circuits are operated at 1 MHz. The FWR presented in [22] is simple and uses one extra-X CCII (EXCCII) and two nMOS switches. A recently reported circuit in [23] uses one double output CCII (DOCCII) along with one rectifying core circuit (RCC). Although it consumes low power, its dynamic range is limited to ± 100μA and operating frequency is 1 MHz. One current differencing buffered amplifier (CDBA) and two diodes are used in [24] to design rectifier and has low operating frequency is 1 MHz. Another Rectifier [25] designed using single dual-X second generation current conveyor (DX-CCII), two diodes and one resistor. It has operating frequency of 5 MHz and dynamic range of 0.3 mA. In [26] one double output current follower (DOCF), four diodes and two resistors are used. It has comparatively low operating frequency. In [27] two current conveyors (CCIIs), and two diodes are used to design rectifier. It also require one auxiliary bias voltage.

In this paper, a new current mode (CM) precision rectifier is proposed using one EXCCII and two nMOS transistors. It can provide + ve FWR output, – ve FWR output, + ve half-wave rectifier (HWR) output, and – ve HWR output, simultaneously without any alteration of topology. It has a very high operating frequency of 125 MHz and above. It is suitable for full cascading. Moreover, it has good dynamic range, low noise, low total harmonic distortion, and very much insensitive to temperature. Monte-Carlo analysis with respect to threshold voltage has also been studied and found to be satisfactory. It does not use any auxiliary bias source.

2 EXCCCII description

A basic EX-CCII normally contains one Y terminal, two X terminals (X1, X2) and two Z terminals (Z1, Z2) along with one terminal for bias current I_0. However, the number of Z terminals may be increased by copying as per requirement. Its port relationships can be characterized by the matrix given in (1). The block diagram of EX-CCII with additional Z terminals is shown in Fig. 1 followed by

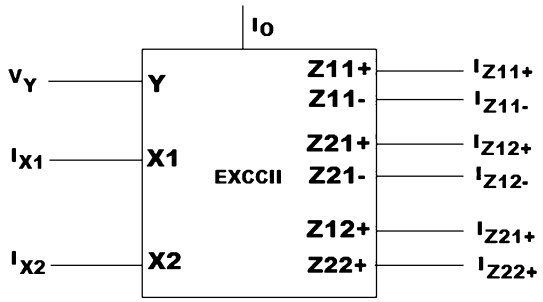


Fig. 1 Block diagram of EX-CCII with additional Z terminals

Fig. 2, where the internal CMOS circuit diagram of it is given.

$$\begin{bmatrix} I_Y \\ V_{X1} \\ V_{X2} \\ I_{Z1\pm} \\ I_{Z2\pm} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{bmatrix} \begin{bmatrix} V_Y \\ I_{X1} \\ I_{X2} \end{bmatrix} \quad (1)$$

3 EXCCII based rectifier circuit and its analysis

The proposed full wave precision rectifier circuit is shown in Fig. 3. It uses one EXCCII building block and two diode-connected- MOSFETs. To understand the operation of the proposed circuit we need to understand the ON and OFF instances of m1 and m2 MOS transistors in sequence to the input current. It is observed by simulation that the voltage developed at the input terminal P when I_{in} is applied as shown in Fig. 4.

The transistor m1 (m2) is ON (OFF) for $I_{in} > 0$ i.e. for positive half cycle & OFF (ON) for $I_{in} < 0$ i.e. for negative half cycle.

Hence, when $I_{in} > 0$.

$I_{X1} = +I_{in}$, hence as per (1).

$$I_{Z11+} = +I_{in};\ I_{Z11-} = -I_{in};\ I_{Z12+} = +I_{in}; \quad (2)$$

Similarly, for $I_{in} < 0$,

$I_{X2} = -I_{in}$, hence from (1).

$$I_{Z21-} = +I_{in};\ I_{Z21+} = -I_{in};\ I_{Z22+} = -I_{in}; \quad (3)$$

Therefore using (3) and (4) we obtain outputs from rectifier of Fig. 3 as follows:

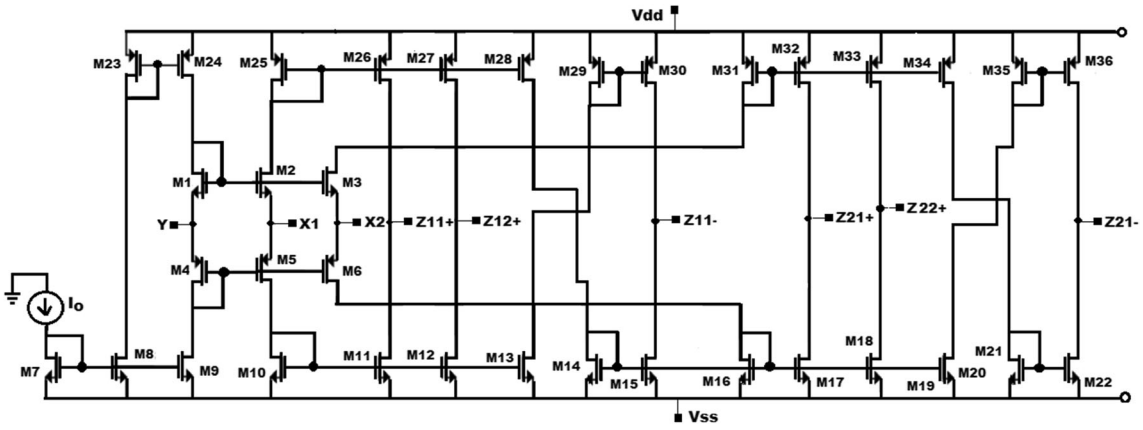


Fig. 2 CMOS circuit diagram of EX-CCII

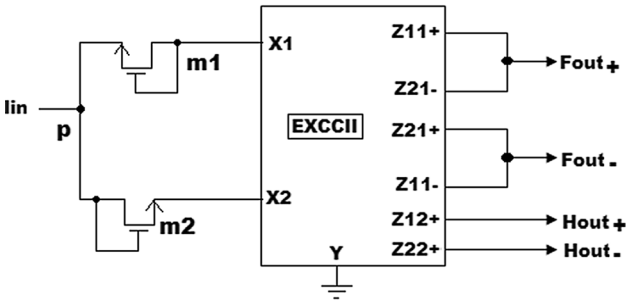


Fig. 3 Proposed full wave and half wave rectifier

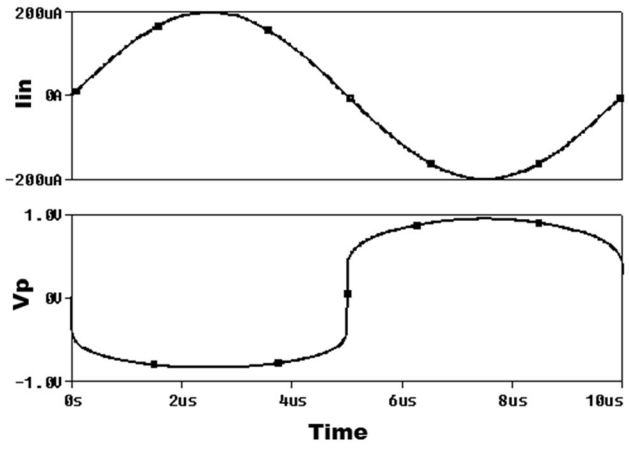


Fig. 4 Input current (I_{in}) and voltage (Vp) variation with time

Table 1 Aspect ratio of various transistors of EX-CCII

Transistors	Aspect ratio W(μm)/L(μm)
M_1,M_2,M_3	7.2/0.36
M_4,M_5,M_6	11.52/0.36
All other NMOS	4.32/0.36
All other PMOS	7.2/0.36

It may be observed that we can get four types of rectifier implemented such as positive-full-wave-rectifier (+ FWR), negative-full-wave-rectifier (− FWR), positive-half-wave-rectifier (+ HWR), and negative-half-wave-rectifier(− HWR) simultaneously from the terminals F_{out} + , F_{out}-, H_{out} + , and H_{out}-, respectively.

4 Simulation result

The circuit is simulated using Cadence ORCAD PSPICE simulator in 0.18 μm TSMC CMOS technology process parameters. Supply voltage and bias current are taken as ± 1.25 V and I_B = 5 μA, respectively. The W/L ratio of the various transistors of EX-CCII is shown in Table 1. The

$$Positive\ Full\ -\ \text{wave}\ rectified\ output = F_{out+} = \begin{cases} I_{Z11+} = +I_{in}, & when\ I_{in} > 0 \\ I_{Z21-} = +I_{in}, & when\ I_{in} < 0 \end{cases}$$

$$\text{Negative Full} - \text{wave rectified output} = F_{out-} = \begin{cases} I_{Z11-} = -I_{in}, & \text{when } I_{in} > 0 \\ I_{Z21+} = -I_{in}, & \text{when } I_{in} < 0 \end{cases}$$

$$\text{Positive Half} - \text{wave rectified output} = H_{out+} = I_{Z12+} = +I_{in}, \quad \text{when } I_{in} > 0$$

$$\text{Negative Half} - \text{wave rectified output} = H_{out-} = I_{Z22+} = -I_{in}, \quad \text{when } I_{in} > 0$$

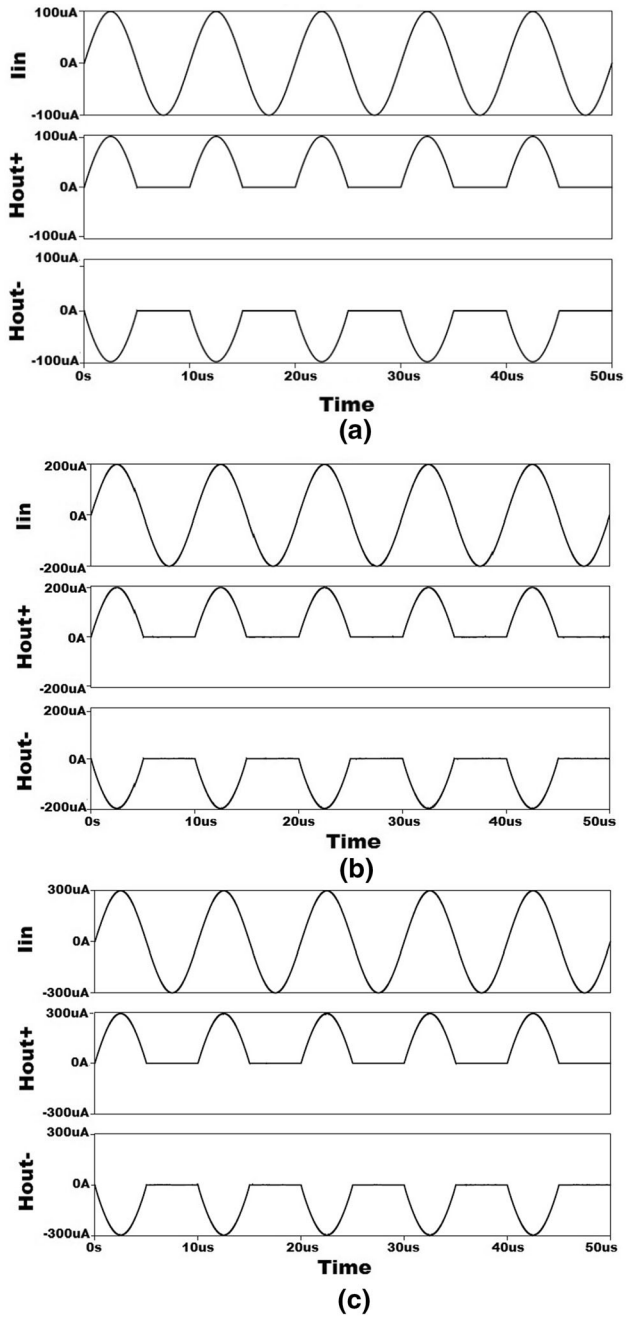


Fig. 5 Simulated rectified outputs of HWR in dual phase for input currents of **a** 100 μA **b** 200 μA **c** 300 μA at 100 kHz

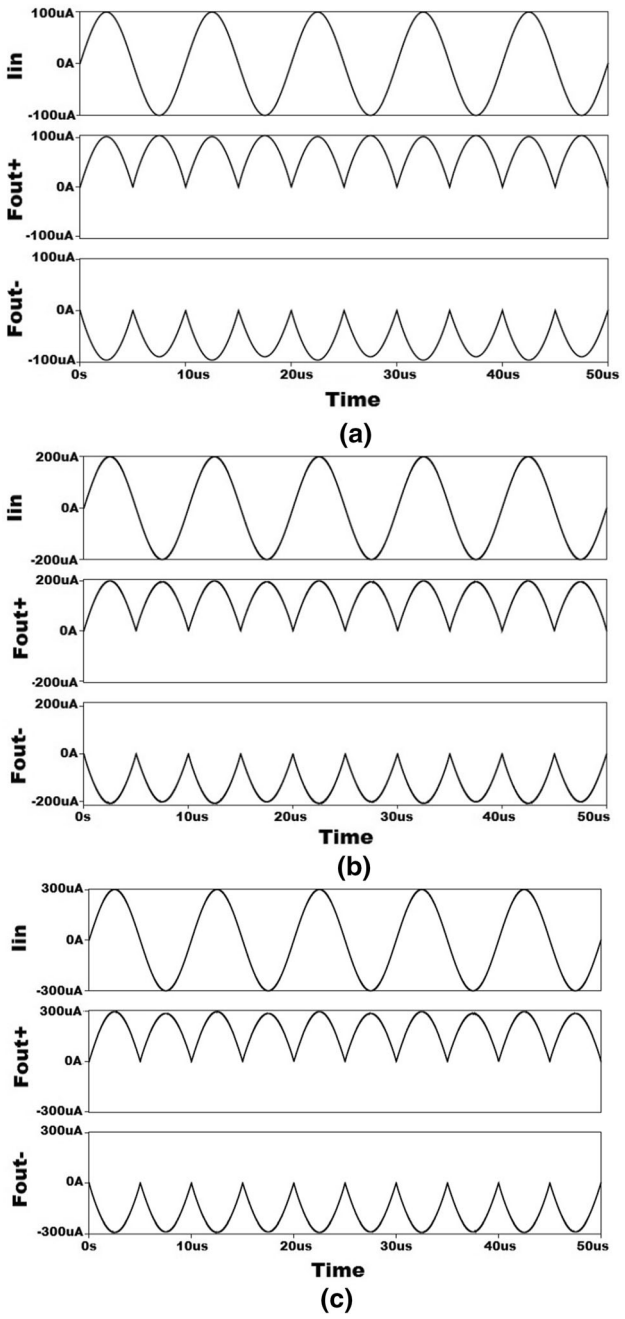


Fig. 6 Simulated rectified outputs of FWR in dual phase for input currents of **a**100 μA **b** 200 μA **c** 300 μA at 100 kHz

W/L ratio of the two MOS transistors (m1 and m2) are taken as $(W/L)_{m1} = (W/L)_{m2} = (0.8\ \mu m / 0.18\ \mu m)$.

The simulated rectifier responses for sinusoidal input currents having amplitude of 100 μA, 200 μA, and 300 μA at 100 kHz are shown in Fig. 5 and Fig. 6 for half-wave (HWR) and full-wave rectifiers (FWR), respectively. The proposed rectifier is also tested for the frequencies of 1 MHz, 10 MHz, 50 MHz, 100 MHz, and 125 MHz at a current of 200 μA for FWR as is shown in Fig. 7. It can even work for higher frequency. The DC transfer characteristics of the rectifier have also been performed. The corresponding transfer characteristics of the positive and negative HWR and the FWR are shown in Fig. 8 and Fig. 9, respectively. It is observed that its input dynamic range is ± 500 μA. Temperature analysis is also performed for the HWR and FWR with the application of the sinusoidal input signal of 200 μA, 100 kHz at the temperatures of 25 °C, 50 °C, and 75 °C. The simulated outputs are

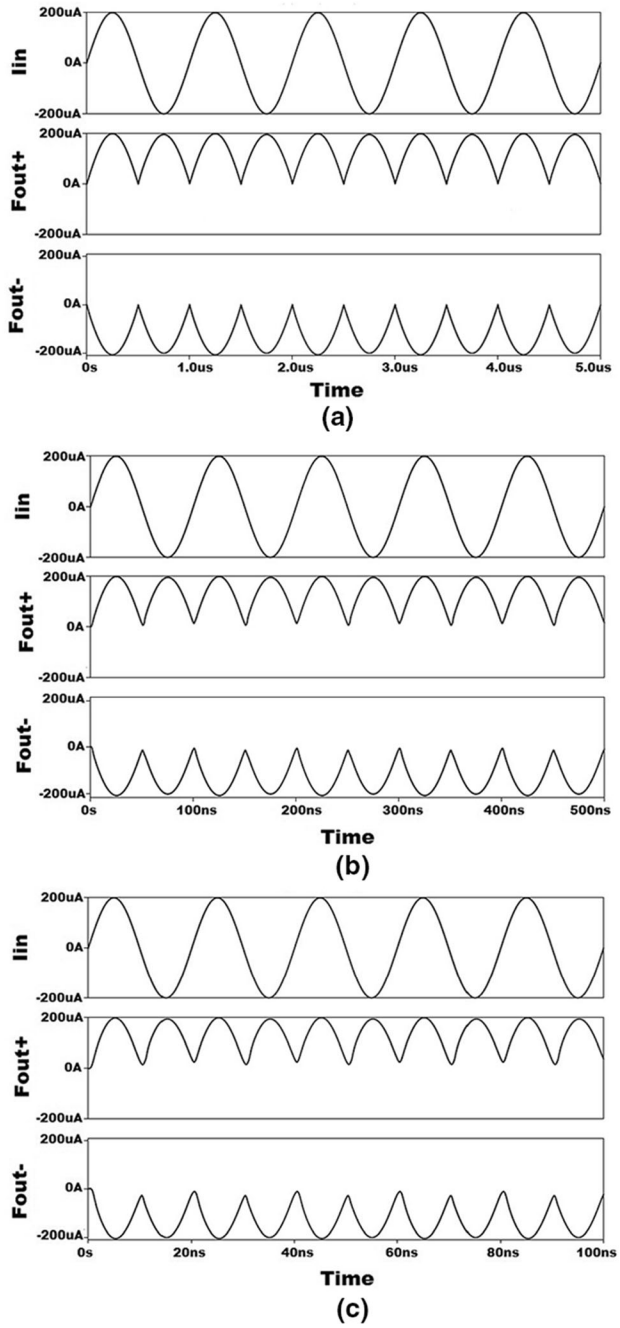


Fig. 7 Simulated rectified outputs of FWR at a different frequency in dual phase for input current of 200 μA for **a** 1 MHz **b** 10 MHz **c** 50 MHz **d** 100 MHz **e** 125 MHz

shown in Fig. 10. It is evident that the responses are almost insensitive to temperature variation.

To obtain the rectifiers accuracy DC value transfer (ρ_{DC}) and RMS error (ρ_{RMS}) are analyzed. They are mathematically defined as:

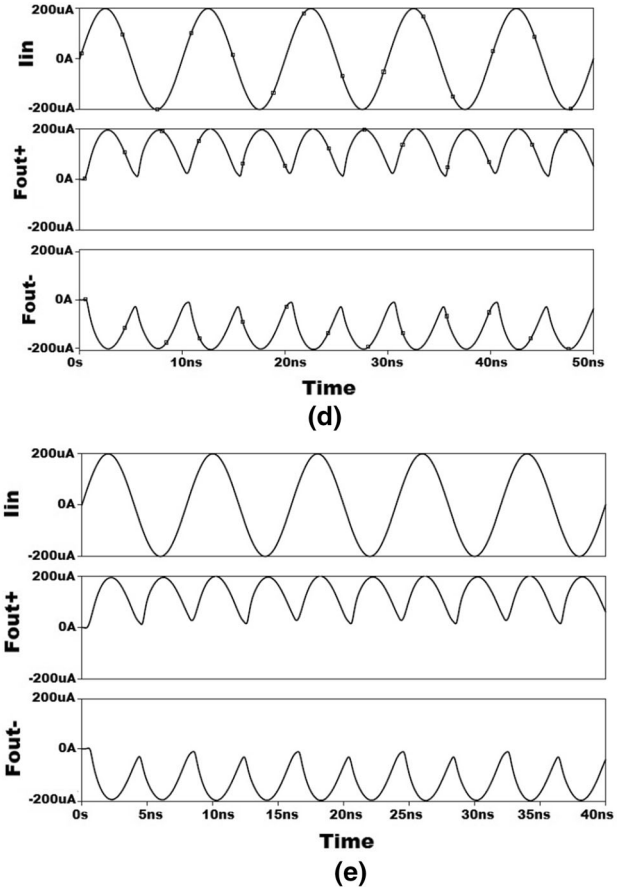


Fig. 7 continued

$$\rho_{DC} = \frac{\int_T I_{oa}(t)dt}{\int_T I_{oi}(t)dt} \tag{4}$$

$$\rho_{RMS} = \sqrt{\frac{\int_T [I_{oa}(t) - I_{oi}(t)]^2 \, dt}{\int_T I_{oi}^2(t) \, dt}} \tag{5}$$

where I_{oa} is the actual value of output and I_{oi} is the ideal value of output. Ideally the value of $\rho_{DC} = 1$, and $\rho_{RMS} = 0$ Fig. 11 shows the simulated ρ_{DC} vs frequency and Fig. 12 shows ρ_{RMS} vs frequency. It is evident from Fig. 11 and Fig. 12 that the proposed rectifier can be used for very high frequency, which is in harmony with the result obtained in Fig. 7.

The average value of output current for the FWR is shown in Fig. 13 and its value is ± 126.5 μA, which is closed to the theoretical value of $2\left(\frac{I_m}{\pi}\right)$ = ± 127.32 μA, where I_m = 200 μA.

Total harmonic distortion (THD) with respect to frequency is shown in Fig. 14. It is observed that for an input signal of 200 μA the THD is approximately − 7.6 dB to − 8 dB. The output noise vs frequency is plotted in Fig. 15. It is observed that the output noise of the proposed circuit is very low. Moreover, the effect of threshold

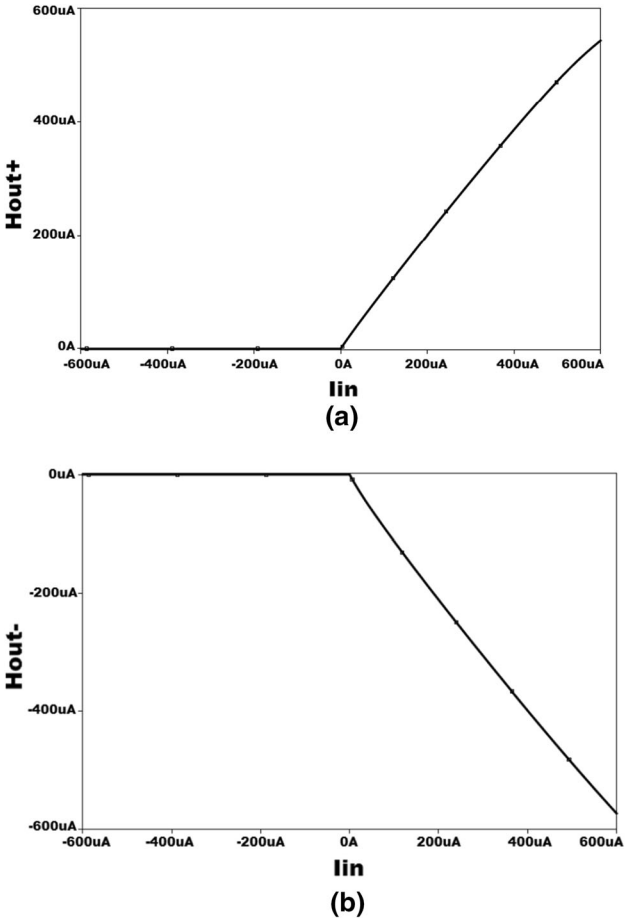


Fig. 8 Simulated half-wave-rectifier DC transfer characteristics in dual phase **a** positive **b** negative

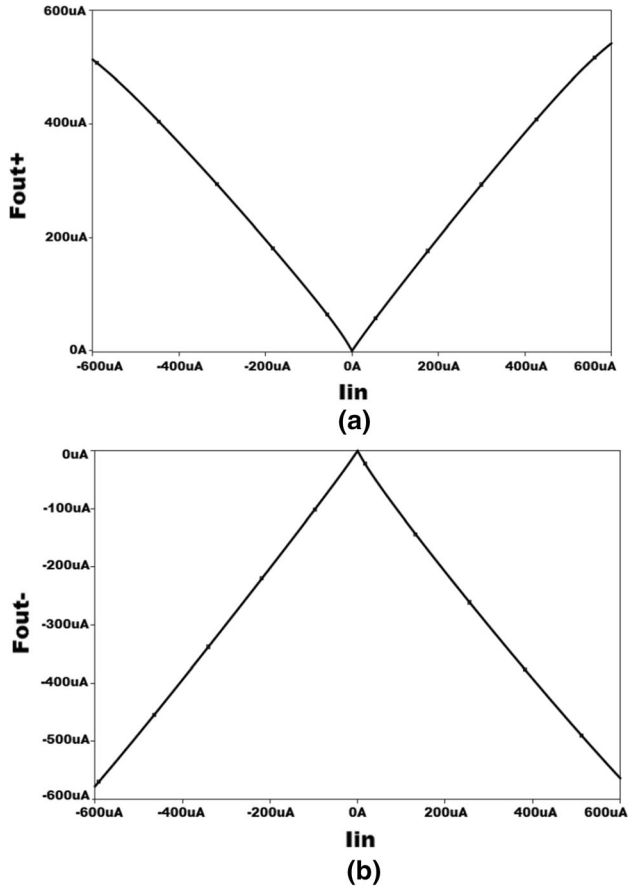


Fig. 9 Simulated full-wave-rectifier DC transfer characteristics in dual phase **a** positive **b** negative

voltage on rectifier output is studied by Monte Carlo analysis. The Monte Carlo simulation of the proposed circuit for the 1% deviation of the threshold voltage of all the MOS transistors for 50 runs is shown in Fig. 16. It is observed that the effect is not significant.

5 Application

The rectifier has various applications. One of the applications of the proposed rectifier is to obtain the root mean square (RMS) value for a sinusoidal current signal. In this application the average of the signal is obtained first and then the RMS is obtained. Average signal magnitude is defined by

$$I_{avg} = \frac{1}{T}\int_0^T |I(t)|dt \tag{6}$$

here I(t) = AC signal, T = time period of the AC signal, and I_{avg} = rectified average value of I(t). To obtain DC component, AC signal is first rectified and then passes through a low-pass (LP) filter. Let $I(t) = I_m \sin(2\pi ft)$, where I_m = peak value of the signal, and f = 1/T = frequency of the signal. The average value of I(t) is computed to be

$$I_{avg} = \frac{2}{\pi} I_m = 0.637 I_m \tag{7}$$

The expression to compute RMS value is given as

$$I_{RMS} = \sqrt{\frac{1}{T}\int_0^T I^2(t)dt} \tag{8}$$

The RMS value for sinusoidal I(t) is

$$I_{RMS} = \frac{1}{\sqrt{2}} I_m = 0.707 I_m \tag{9}$$

From (7) and (8) we get

$$\frac{I_{RMS}}{I_{avg}} = \frac{\pi}{2\sqrt{2}} \approx 1.11 \tag{10}$$

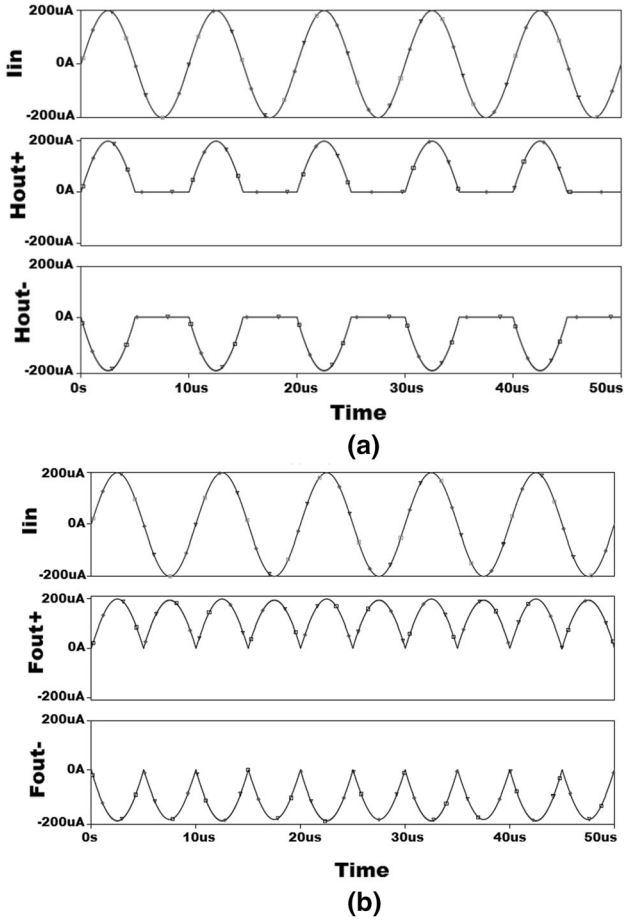


Fig. 10 Simulated responses at different temperatures(square 25 °C diamond 50 °C triangle 75 °C) for **a** HWR **b** FWR

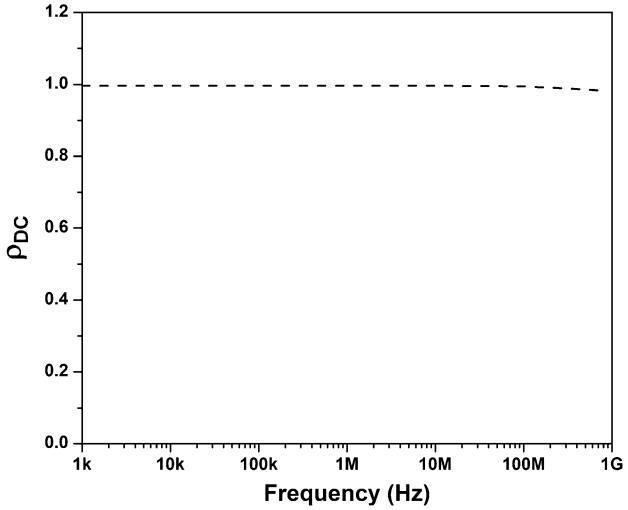


Fig. 11 Variation of ρ_{DC} with frequency for FWR of Fig. 3

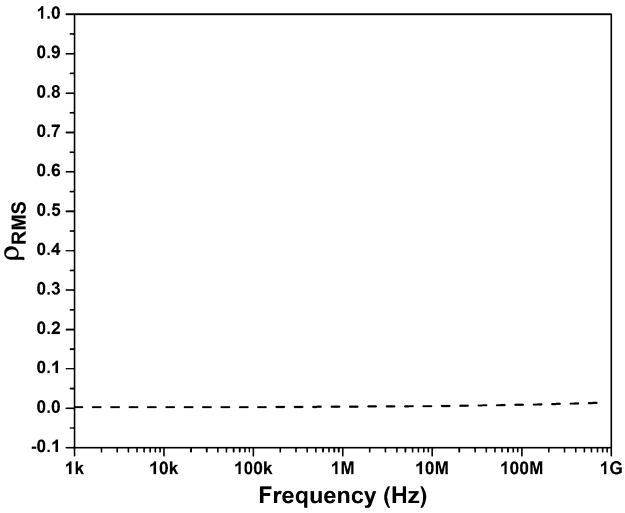


Fig. 12 Variation of ρ_{RMS} with frequency for FWR of Fig. 3

In line with [33], the average and RMS values of the sinusoidal current signal may be obtained using the proposed full wave rectifier as shown in Fig. 17.

In case of current mode signal a low-pass filter is obtained by properly choosing the value of parameters of m3 MOS transistor and value of C. Output current of the rectifier (I_{out}) is applied to the input of the filter in which the ac components of the signal is filtered by the capacitor C the DC component is entered into m3 and then it is copied by the current mirror transistors m4 and m5. By selecting m3 and m4 transistors identical I_{avg} current can be obtained at m4. Using (10) I_{RMS} is obtained at the drain of m5 by considering $g_{m5}/g_{m3} = 1.11$, where gm is transconductance of the corresponding transistor.

The capacitor value should be such a large value that the ripple remains within the considerable range. In other words,

$$C >> \frac{g_{max}}{4\pi f_{min}} \tag{11}$$

here f_{min} = lowest frequency of interest, and g_{max} = highest value of transconductance, that is at I_m.

5.1 Simulation result

For the simulation of the circuit of Fig. 17, the supply voltage and bias current are taken as ± 1.25 V and I_B = 10 μA, respectively for EX-CCII. The W/L ratio of the various transistors of EX-CCII and MOS transistors (m1 and m2) are taken as before. The aspect ratios for m3, m4, and m5 MOS transistors are taken as (23 μm/2 μm), (23 μm/2 μm), and (25.53 μm/2 μm) respectively. The capacitor (C) value is taken as 4nF. A sinusoidal current signal of 50 μA, 500 kHz is applied to the circuit. The simulation

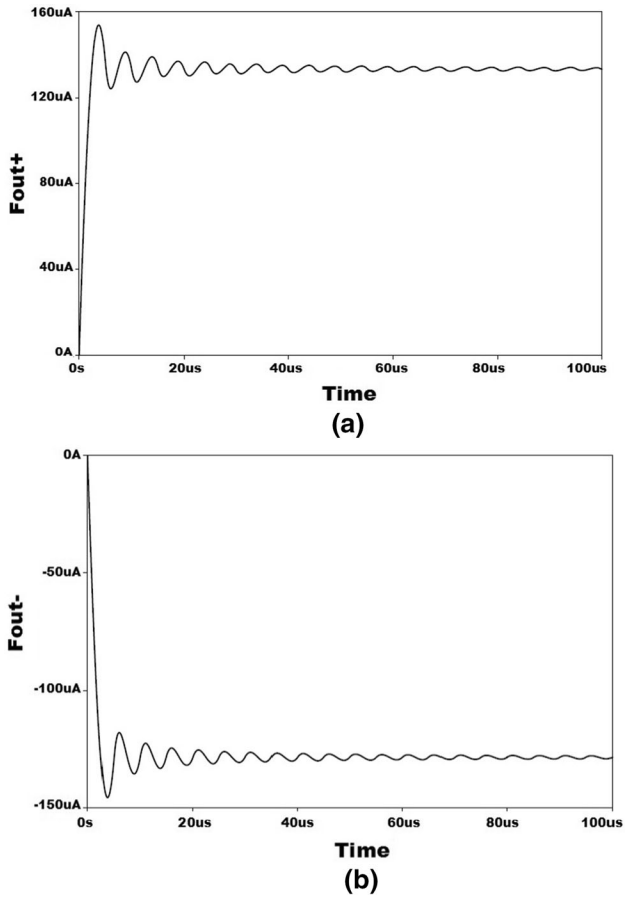


Fig. 13 Average DC current output of FWR for dual phase **a** positive **b** negative

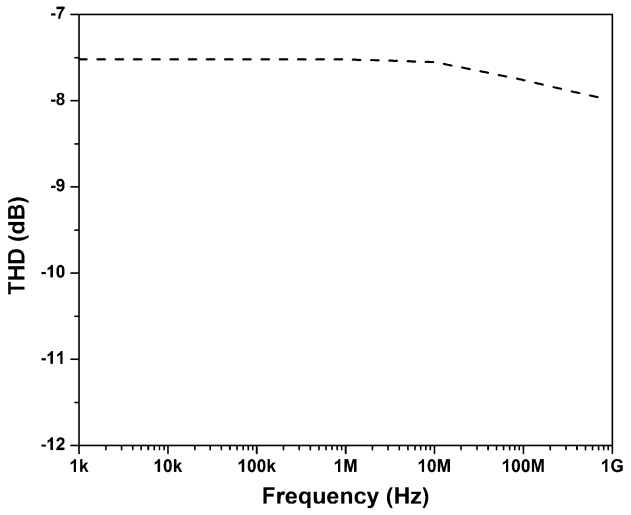


Fig. 14 Variation of THD of FWR output as a function of frequency

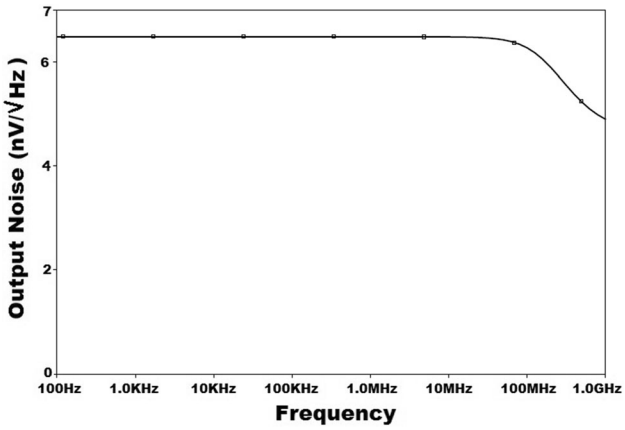


Fig. 15 Variation of output noise of FWR output with frequency

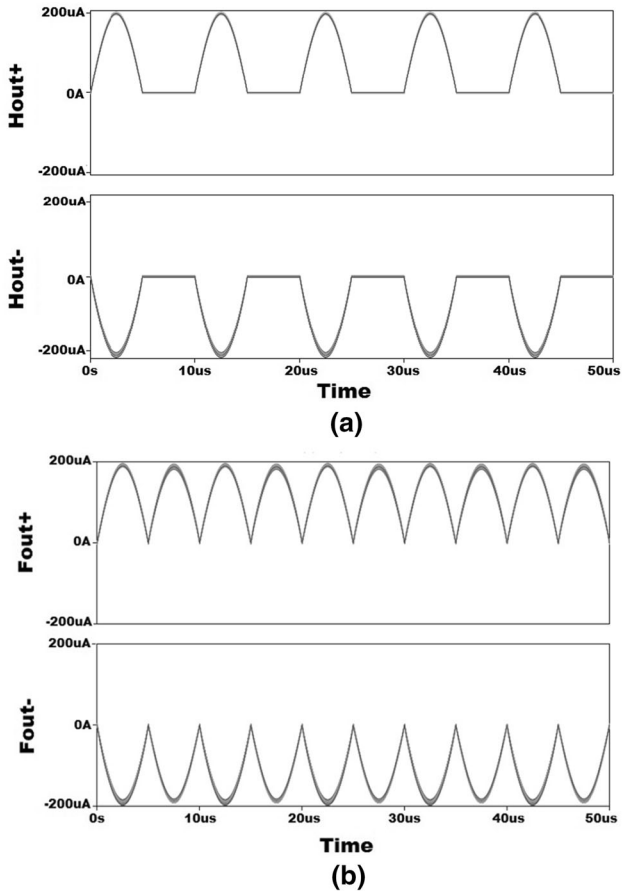


Fig. 16 Monte Carlo analysis of positive and negative **a** HWR and **b** FWR with 1% deviation of the threshold voltage of all MOS transistors for 50 runs

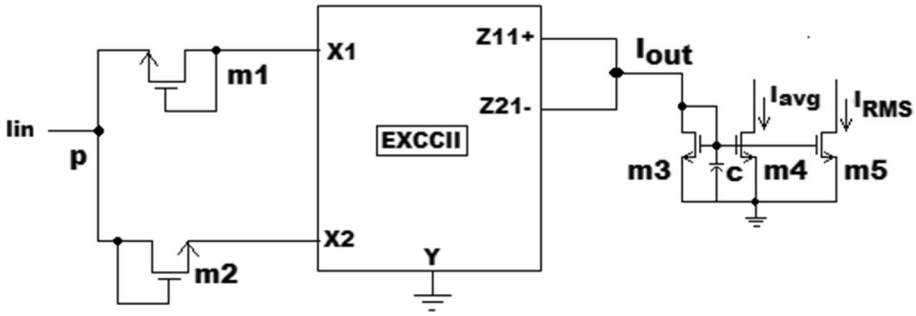


Fig. 17 Fullwave rectifier based RMS value computation for sinusoidal signal

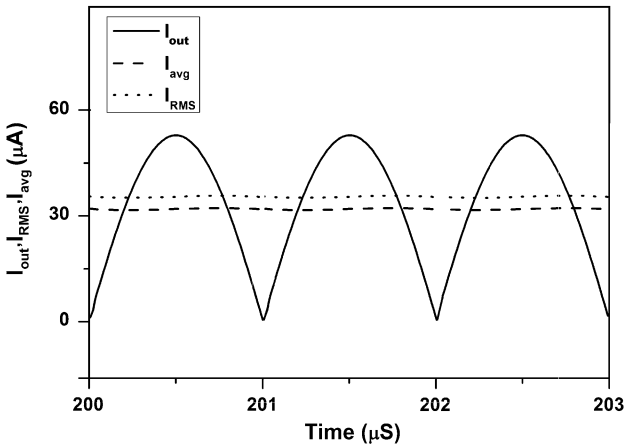


Fig. 18 Simulated I_{out} (rectified), I_{avg}, and I_{RMS}, for sinusoidal input with 50 μA, 500 kHz

results are shown in Fig. 18. The I_{avg} and I_{RMS} are obtained as 31.80 μA, and 35.30 μA respectively, as expected, with a ripple of around 0.9%.

6 Comparison with the available CM rectifiers

A comparison of the proposed work with the similar type of available current mode (CM) work is given in Table.2. It is observed that the available circuits have one or more of the shortcomings enumerated as follows: use of two or more analog building blocks in the count [13, 14, 17, 23, 27], use of resistor(s) [11, 16, 17, 21, 25, 26], use of excessive number of diodes in the count [11, 16, 17, 24–27], not fully cascadable [13, 14, 21, 24], comparatively low operating frequency [11–18, 21–23], need of auxiliary bias current/voltage [11, 14, 17, 20], not dual phase rectified output [11–22], higher power consumption [11, 18, 20, present work], high output noise [21], comparatively low input dynamic range [11, 21–27], and high THD [18]. It may be noted that the present work will consume lesser power as well as need lesser number of MOS transistors in the count for implementation of EXCCII if the circuit is operated in single phase instead of dual phase.

7 Conclusion

In this work, a new current-mode (CM) precision rectifier has been presented. The proposed circuit is designed using a single EXCCII analog building block and two n-MOS transistors. It can provide rectified outputs in dual phase simultaneously from the same topology. It possesses low input and high output impedance, hence suitable for full cascading. The circuit is useful for very high frequency; 125 MHz and above. Moreover, fairly good performances are observed for other important parameters such as power consumption, output noise, input dynamic range, temperature stability, total harmonic distortion, and Monte Carlo analysis. PSPICE simulation result with 0.18 μm technology agrees well with the theoretical proposition. The

Table 2 Comparison with the previously published CM rectifiers

Ref	Analog building block (ABB)	No of transistor	No of resistor used	No of diode/ additional MOS	Input/ output Imp.	Max freq. shown (MHz)	Supply vol. (Volt)	Auxiliary bias current/ voltage used	Rectifier o/p type	Power consume (mW)	O/p noise in $nV/\sqrt{Hz}$	Tech. used (μm)	I/p dyna-mic range (mA)	Temp. stability	THD
[11]	1-CDTA	26	1	4/0	L/H	5	± 1.8	Yes	Single phase	6.31	NA	0.35	± 0.21	NA	NA
[12]	1-CDTA	4	0	2/0	L/H	5	NA	No	Single phase	NA	NA	NA	NA	NA	NA
[13]	1-CCII + 1-UVC	NA	0	2/0	L/L	0.1	NA	No	Single phase	NA	NA	NA	NA	NA	NA
[14]	1-CCII + 1-UVC	NA	0	2/0	L/L	1	NA	Yes	Single phase	NA	NA	NA	NA	NA	NA
	2-CCII	NA	0	2/0	L/L	1	NA	Yes	Single phase						
[16]	1-CDTA	NA	1	4/0	L/H	10	NA	No	Single phase	NA	NA	NA	± 3	NA	NA
[17]	2-CDTA 1-MCF	24	2	4/0	L/H	1	± 3	Yes	Single phase	NA	NA	BJT	NA	NA	NA
[18]	1-MZC-CDTA	28	0	0/2	L/H	0.01	± 1.8	No	Single phase	14	NA	0.35	± 0.3	NA	22% upto 300 uA
[20]	1-CDTA	21	0	0/2	L/H	100	± 1.5	Yes	Single phase	1.12	3.52	0.18	± 2[a]	NA	NA
[21]	1-OTA	8	2	2/0	H/H	1	± 1.5	No	Single phase	NA	12	0.25	± 0.2	Yes	NA
	1-DVCC	12	3	2/0	H/H	1	± 1.5	No	Single phase	NA	20	0.25	± 0.2	Yes	NA
[22]	1-EXCCII	27	0	0/2	L/H	10	± 1.25	No	Single phase	0.3	6.4	0.25	± 0.2	Yes	− 10 dB
[23]	1-DOCCII 1-RCC	34	0	0/0	L/H	1	± 0.9	No	Dual phase	0.188	NA	0.18	± 0.1	NA	− 12.8 dB
[24]	1-CDBA	20	0	2/0	H/L	1	± 1.5	No	Single phase	NA	NA	0.25	NA	NA	NA
[25]	1-DXCCII	20	1	2/0	L/H	5	NA	NA	Single phase	NA	NA	0.35	± 0.3	NA	NA
[26]	1-DOCF	20	2	4/0	L/H	1	NA	NA	Single phase	NA	NA	NA	NA	NA	NA
[27]	2-CCII	30	0	2/0	L/H	5	± 1.5	Yes	Single phase	NA	NA	0.35	NA	NA	NA

Table 2 (continued)

Ref	Analog building block (ABB)	No of transistor	No of resistor used	No of diode/ additional MOS	Input/ output Imp.	Max freq. shown (MHz)	Supply vol. (Volt)	Auxiliary bias current/ voltage used	Rectifier o/p type	Power consume (mW)	O/p noise in $nV/\sqrt{Hz}$	Tech. used (µm)	I/p dynamic range (mA)	Temp. stability	THD
Our work	1-EXCCII	36	0	0/2	L/H	125	± 1.25	No	Dual-phase	0.62[b]	6.5	0.18	± 0.5	Yes	− 7.52 dB

Single phase = Either + ve or – ve half-wave & full-wave rectifier output; Dual phase = both + ve and – ve half-wave and full-wave rectifier output

NA Not available, *MCF* Multi-output current follower, *RCC* Rectifying core circuit

[a]I/p and o/p currents are unequal

[b]For single phase it will be lesser

sinusoidal to RMS conversion is also shown over here as an application of the proposed rectifier.

References

1. Maheshwari, S. (2018). Voltage-mode full-wave precision rectifier and an extended application as ASK/BPSK circuit using a single EXCCII. *International Journal of Electronics and Communications, 84*, 234–241.
2. Maheshwari, S. (2017). Realization of simple electronic functions using EXCCII. *Journal of Circuits Systems and Computers, 26*(11), 1–12.
3. Yuce, E., Minaei, S., & Ibrahim, M. A. (2017). A novel full-wave rectifier/sinusoidal frequency doubler topology based on CFOAs. *Analog Integrated Circuits and Signal Processing, 93*(2), 351–362.
4. Ibrahim, M. A., Yuce, E., & Minaei, S. (2016). A new DVCC-based fully cascadable voltage-mode full-wave rectifier. *Journal of Computational Electronics, 15*(4), 1440–1449.
5. Koton, J., Vrba, K., & Herencsar, N. (2014). Voltage-mode full-wave rectifier based on DXCCII. *Analog Integrated Circuits and Signal Processing, 81*(1), 99–107.
6. Beg, P. I., Khan, A., & Maheshwari, S. (2012). Biphase amplifier based precision rectifiers using current conveyors. *International Journal of Computer Applications, 42*(3), 14–18.
7. Koton, J., Herencsar, N., & Vrba, K. (2010). Minimal configuration precision full-wave rectifier using current and voltage conveyors. *IEICE Electronics Express, 7*(12), 844–849.
8. Kumngem, M., Saengthong, P., & Junnapiya, S. (2009). DDCC-based full-wave rectifier. *International colloquium on signal processing & its applications*, pp. 312–315.
9. Djukic, S. R. (2008). Full-wave current conveyor precision rectifier. *Serbian Journal of Electrical Engineering, 5*(2), 263–271.
10. Anuntahirunrat, K., Tangsrirat, W., Riewruja, V., & Surakampontorn, W. (2004). Sinusoidal frequency doubler and full-wave rectifier based on translinear current-controlled current conveyors. *International Journal Electronics and Communications, 91*(4), 227–239.
11. Biolek, D., Hancioglu, E., & Keskin, A. U. (2008). High-performance current differencing transconductance amplifier and its application in precision current-mode rectification. *International Journal of Electronics and Communications, 62*(2), 92–96.
12. Khateb, F., Vavra, J., & Biolek, D. (2010). A novel current-mode full wave rectifier based on one CDTA and two diodes. *Radio engineering, 19*(3), 437–445.
13. Koton, J., Herencsár, N., Vrba, K., & Cicekoglu, O. (2010). Versatile precision full-wave rectifier using current and voltage conveyor. *IEEE international conference on applied electronics*, pp. 79–82.
14. Koton, J., Herencsar, N., & Vrba, K. (2011). Current and voltage conveyors in current and voltage mode precision full-wave rectifiers. *Radio engineering, 20*(1), 19–24.
15. Koton, J., Lahiri, A., Herencsar, N., & Vrba, K. (2011). Current-mode dual-phase precision full-wave rectifier using current-mode two-cell winner-takes-all (WTA) circuit. *Radio engineering, 20*(2), 428–432.
16. Koton, J., Herencsar, N., Vrba, K., & Minaei, S. (2011). Precision full-wave current-mode rectifier using current differencing transconductance amplifier. *IEEE international conference on communication software and networks (ICCSN)*, pp. 460–463.

17. Tangsriirat, W., Pukkalanun, T., & Surakampontorn, W. (2011). Synthesis of current differencing transconductance amplifier-based current limiters and its applications. *Journal of Circuits Systems and Computers, 20*(2), 185–206.
18. N Pandey R Pandey 2013 Current mode full-wave rectifier based on a single MZC-CDTA. Active and Passive Electronic Components, Article ID 967057, pp. 1–5.
19. Shaterian, M., Twigg, C. M., & Azhari, J. (2013). An MTL-based configurable block for current-mode nonlinear analog computation. *IEEE Transactions on Circuits Systems II Express Briefs, 60*(9), 587–591.
20. Kacar, F., & Basak, M. E. (2014). A new mixed mode full-wave rectifier realization with current differencing transconductance amplifier. *Journal of Circuits, Systems, and Computers, 23*(7), 1–15.
21. Sagbas, M., Minaei, S., & Ayten, U. E. (2016). Component reduced current-mode full-wave rectifier circuits using a single active component. *IET Circuits Devices & Systems, 10*(1), 1–11.
22. Agrawal, D., & Maheshwari, S. (2017). Current mode precision full wave rectifier circuits. *Circuits Systems Signal Process*. https://doi.org/10.1007/s00034-017-0531-8.
23. Safari, L., Yuce, E., & Minaei, S. (2018). A new low-power current-mode MOS only versatile precision rectifier. *International Journal of Electronics and Communications, 83*, 40–51.
24. Erkan, S., Ayten, U. E., & Sagbas, M. (2015). Current-mode full-wave rectifier circuits using current differencing buffered amplifier. *38th international conference on telecommunications and signal processing (TSP)*, pp. 344–348.
25. Koton, J., Herencsar, N., Vrba, K. (2011). Current-mode precision full-wave rectifier using single DXCCII and two diodes. *Circuit theory and design (ECCTD), 20th European conference.*https://doi.org/10.1109/ECCTD.2011.6043400, 2011, pp. 508–511.
26. Tiliute, D. E. (2003). Full-wave current-mode precision rectifiers using unity-gain cells. *Elektronika Ir Elektrotechnika, 49*(7), 26–29.
27. Koton, J., Herencsar, N., & Vrba, K. (2010). Minimal configuration precision full-wave rectifier using current and voltage conveyors. *IEICE Electron Express, 7*, 844–849.
28. Yuce, E., & Alpaslan, H. (2012). A CMOS current rectifier configuration suitable for integration. *Journal of Circuits, Systems, and Computers, 21*(7), 12.
29. Minaei, S., & Yuce, E. (2010). New squarer circuits and a current-mode full-wave rectifier topology suitable for integration. *Radio engineering, 19*(4), 657–661.
30. Chang, C. C., & Liu, S. I. (2000). Current-mode full-wave rectifier and vector summation circuit. *Electronics Letters, 36*(19), 1599–1600.
31. Burapattanasiri, B. (2009). High-precision multi-wave rectifier circuit operating in low voltage + 1.5 Volt current mode. *International Journal of Computer Science and Information Security, 6*(3), 160–164.
32. Petrovic, P. B. (2017). Variable mode CMOS full-wave rectifier. *Analog Integrated Circuits and Signal Processing, 90*, 659–668.
33. Wang, Z. (1990). Novel pseudo RMS current converter for sinusoidal signals using a CMOS precision current rectifier. *IEEE Transactions on Instrumentation and Measurement, 39*(4), 670–671.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Rupam Das received his B.Tech. Degree in Electronics & Communication Engineering in 2006 and M.Tech. Degree in VLSI & Microelectronics in 2008 from West Bengal University of Technology, West Bengal, India. He is a part time research scholar in Indian Institute of Technology (Indian School of Mines), Dhanbad, India. Currently he is working as an Assistant Professor in Electronics and Communication Engineering at Asansol Engineering College, Asansol, India. His main research interest is in the area of current mode analog integrated circuit design.



Sajal K. Paul did his B.Tech., M.Tech., and Ph.D. in Radio Physics and Electronics from the Institute of Radio Physics and Electronics, University of Calcutta. He has served Webel Telecommunication Industries, Kolkata; Indira Gandhi National Open University (IGNOU), Kolkata; Advanced Training Institute for Electronics & Process Instrumentation (ATI-EPI), Hyderabad; North Eastern Regional Institute of Science & Technology (NERIST), Nirjuli and Delhi College of Engineering (DCE), Delhi in various capacities. He has served the Indian Institute of Technology (Indian School of Mines), Dhanbad, India as Dean (Faculty), Head of the Department of Electronics Engineering, and at present Professor of the same department. His research interest includes Microelectronic Devices, Electronic Properties of Semiconductor and Bipolar and MOS Analog Integrated Circuits. Dr. Paul has more than 115 research publications in International and National journals of repute and conferences.

IET Power Electronics

Research Article

IET Journals
The Institution of Engineering and Technology

Resonant controller of single-stage off-board EV charger in G2V and V2G modes

ISSN 1755-4535
Received on 30th September 2019
Revised 26th November 2019
Accepted on 19th December 2019
E-First on 19th February 2020
doi: 10.1049/iet-pel.2019.1111
www.ietdl.org

Aakash Kumar Seth[1], *Mukhtiar Singh*[1] ✉

[1]*Electrical Engineering, Delhi Technological University, Bawana Road, New Delhi, India*

✉ *E-mail: smukhtiar_79@yahoo.co.in*

Abstract: This study presents the control design and implementation of a three-phase single-stage off-board electric vehicle (EV) charger. The charger is controlled to perform four quadrants operation in active–reactive (*P*–*Q*) power plane. Generally, the control of the EV charger consists of two control loops, i.e. exterior (slow) loop and fast-acting interior loop. Since the signals to be regulated in outer loop are DC in nature, they can be easily handled by regular proportional plus integral regulators. The inner loop control may also be performed in the DC domain, but inaccuracy involved in determining the parameter-dependent decoupling terms and difficulty in setting the gain for fast-acting inner loops make it unfavourable. Therefore, the inner control is performed in AC domain, where the signals are periodic. The proportional plus resonant controller is found to be very useful in tracking such type of periodic signals of fixed frequency, and accordingly, it has been designed and successfully implemented for the proposed work. The proposed methodology has been tested for eight different modes in the *P*–*Q*-plane. The system is simulated in MATLAB/Simulink environment and implemented in real time with a scaled-down hardware prototype.

1 Introduction

The electric vehicles (EVs) demand is predicted to rise exponentially in upcoming years as they appear on the promising alternative of conventional internal combustion engine vehicle [1–4]. EVs are more environmentally friendly and reduce carbon footprint, and thus, provide a better solution in comparison with fuel-based vehicles. The e-vehicular technology that can be interfaced with the grid is known as gridable vehicles. Such vehicles include battery-driven vehicles and plug-in hybrid EVs, and they are in tandem termed as plug-in EVs [5]. Although the industry of such vehicle is drastically increasing, the obstacle of the short driving range is creating a hindrance in the pathway for overruling the automotive industry, where fuel-based conventional vehicles are still a priority for consumers [6]. To overcome the problem of short driving range in EV, the establishment of fast off-board EV chargers is mandatory [7]. Owing to the improvement in battery technology, they can be charged at high current value, tendering better and fast-charging solutions [8–11].

To provide a solution for charging problems, surplus literature is presented [8, 12–17]. The single-stage off-board EV charger consists of a three-phase AC–DC converter. The EV charger attains the requisite voltage level by two-stage conversation in accordance with battery parameters. The topology used in off-board EV charger may be unidirectional or bidirectional [8]. A unidirectional EV charger requires less components and it is easy to control, but they fail to provide power back to the grid from vehicle and reactive power operation [12]. The bidirectional EV charger can perform these operations and can also minimise the adverse effect of EV charging, especially in the low-voltage distribution network by improving system reliability [13–15]. The system reliability maybe further strengthened with the enormous numbers of EVs connected to the electrical network, where they can be controlled to supply or absorb active as well as reactive power if required. With the energy storage capacity, EV acts as the distribution energy storage with the ability to act as a source or sink during grid-to-vehicle (G2V) and vehicle-to-grid (V2G) operating modes, respectively, and also compensate reactive power [16, 17]. Therefore, the smart and coordinated control EV charging infrastructure may enhance the system reliability. However, the inadequate control of large-scale off-board EV chargers may also adversely affect the system.

The control algorithm of a single-stage off-board EV charger comprises an interior and exterior loop with the outer loop having sluggish dynamics in comparison with the interior loop. The control over current can be done either in AC or DC form. Proportional plus integral (PI) controllers are conventionally used techniques for control over DC quantities [7, 18, 19] with a simple design, easy implementation and satisfactory results. The decoupling elements and accurate *dq* transformation [20], however, are required for optimum operation of PI controller with its bandwidth inversely related to the interfacing inductor affecting the tracking capability of the regulator [21]. Increasing the bandwidth requires reducing the inductor value; thereby, increasing the switching harmonics of the AC side and corresponding ripple on the DC side, which is highly unacceptable. These limitations of the PI regulator and its deficiency to track periodic signal signify the use of the proportional plus resonant (PR) controller in the present case.

Moreover, a direct power control (DPC), model predictive control (MPC) and hysteresis control have also been employed for EV charging [16, 22–25]. Here, the DPC always increases high-power ripples due to the dependence on switching table [22]. MPC may also lead to high-power ripples and large distortion in grid current due to the error between reference vector and selected optimal voltage vector [23].

Furthermore, the oscillating switching frequency may have resonance issues while using hysteresis control [16]. Some control techniques based on artificial intelligence have also been discussed in [24, 25]. These control techniques suffer from the deficiency of computational anxiety for online adaptive tuning.

This paper proposes a strategy for controlling active and reactive powers for deploying an off-board single-stage EV charger. The PI controller is provided at the external loop to maintain the simplicity of control architecture as the quantities such as active and reactive powers are DC signals in nature. However, the inner control is performed in AC domain to avoid the complexity in calculating the decoupling terms. The resultant inner reference currents are sinusoidal AC signals of fixed frequency for which PR maybe easily designed. The PR is simply tuned to provide infinite gain at selected resonant frequency [26, 27]. This enables the smooth control of fast-acting inner current loop with a dismal steady-state error. This approach also results in seamless transfer of operating modes from charging to discharging and vice

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

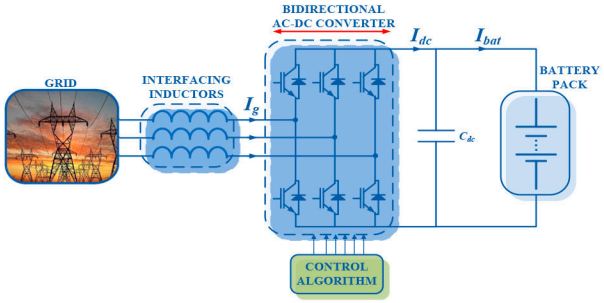


Fig. 1 *Single-stage off-board EV charger*

Table 1 System parameters

Parameters	Symbol	Value (simulation)	Value (experimental)
charger apparent power	S	12.5 kVA	350 VA
grid voltage	V_g	230 V (RMS)	25 V (RMS)
grid frequency	F	50 Hz	50 Hz
interfacing inductor	L_s	4 mH	1 mH
DC-link capacitance	C_{dc}	1100 μF	330 μF
converter switching frequency	f_s	20 kHz	10 kHz
battery voltage	V_{bat}	600 V	100 V

Table 2 Charger operating modes

Mode	Active power, W	Reactive power, VAR	Operation
1	positive	zero	battery charging
2	negative	zero	battery discharging
3	zero	positive	reactive inductive
4	zero	negative	reactive capacitive
5	positive	positive	inductive charging
6	positive	negative	capacitive charging
7	negative	positive	inductive discharging
8	negative	negative	capacitive discharging

versa. Moreover, it also easily handles the dynamic rates of charging and discharging. Practically, it is impossible to implement infinite gain in hardware and the gain needs to be adjusted to certain finite value, which may affect the control performance. The proposed controller is simulated in MATLAB/Simulink under eight different modes of operations. Furthermore, the proposed control logic performance has been validated through a scaled-down hardware prototype experimental setup in real time.

2 System description and operating modes

The system architecture used to investigate the controller deployed for charger as depicted in Fig. 1 with the architecture capable of operating through the four quadrants of the P–Q-plane. Six insulated-gate bipolar transistors (IGBTs) having antiparallel diodes are used in the converter with the blocking capacity of V_{dc} and a maximum conduction current of $\sqrt{2}\ I_g$ with I_g being a grid root-mean-square (RMS) current. The AC side of the converter is interfaced with the grid through inductors (L_s) and a DC-link capacitor (C_{dc}) is connected to the DC side of the converter. The grid-side inductor is used to minimise the ripple content in grid current and helps in attaining sinusoidal current. Limiting the total harmonic distortion below the permissible limit of 5% irrespective of state of charge (SOC) of the battery [28]. This can be attained by appropriate designing of line inductor and grid-side current controller.

The parameters utilised in system are presented in Table 1. Here, the battery nominal voltage is 600 V and its initial SOC is taken at 50% to analyse the charging and discharging behaviour of the battery. The battery voltage depending on the battery SOC and cell configuration maybe above and below the nominal value. In the proposed work, 230 V, 50 Hz supply has been considered, where the minimum DC-link voltage raises up to 563 V, and thus, a high-voltage battery pack is required for single-stage charging. However, in the case of low-voltage battery pack, it is necessary to have second stage of DC–DC converter. Here, it is worth mentioning that the 110 V, 60 Hz supply is more suitable for single-stage charging, where the nominal DC-link voltage is ∼370 V, and hence, eliminates the need of second-stage DC–DC conversion.

The EV charger is controlled to perform four-quadrant operation in eight different modes as listed in Table 2. Here, positive active power means the transfer of power from G2V and vice versa. Similarly, positive reactive power associated with inductive reactive power operation and negative with capacitive reactive power operation. First four modes are associated with individual operation of either active or reactive power and next four modes show the concurrent operation of active and reactive powers. For charging and discharging operations, the current is exactly in phase and out of phase, respectively. Similarly, exact 90° lag behind and lead in case of inductive and capacitive operations, respectively. In the rest four modes, the current is lag/lead, depending on the reactive power operation of the charger.

3 System control development

The proposed EV charger controller is shown in Fig. 2. Here, the main objective of the proposed controller is to track the input active power command (P_{cmd}) and reactive power command (Q_{cmd}) by using regular PI controller. The active power command is directly requested by the user, and is dependent on the battery pack rating, current SOC and user charging choice (i.e. slow, fast or semi-fast). Reactive power can be directly demanded by utility grid. For optimal use of charger, the grid apparent power should be equal to the rating of EV charger. Moreover, to control the grid-

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

side current, sinusoidal reference current is generated, which is further tracked by the PR controller.

For the proposed control system, active and reactive powers are calculated at the grid side in the *dq*-frame. For this, abc to *dq0* transformation, phase-locked loop (PLL) and low-pass filter (LPF) are utilised. An LPF is used to eliminate the 2–*f* component present in measured active (*P*) and reactive (*Q*) powers, whereas PLL is for maintaining the synchronism with the utility grid [29].

For computing *P* and *Q*, the grid current I_{abc} and grid voltages V_{abc} are transformed into the *dq*-frame using Parks transformation [30]

$$\begin{bmatrix} V_{gd} \\ V_{gq} \end{bmatrix} = \frac{2}{3} \times T \times \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} i_{gd} \\ i_{gq} \end{bmatrix} = \frac{2}{3} \times T \times \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} \tag{2}$$

where

$$T = \begin{bmatrix} \cos(\omega t) & \cos\left(\omega t - \frac{2\pi}{3}\right) & \cos\left(\omega t - \frac{4\pi}{3}\right) \\ \sin(\omega t) & \sin\left(\omega t - \frac{2\pi}{3}\right) & \sin\left(\omega t - \frac{4\pi}{3}\right) \end{bmatrix} \tag{3}$$

After transformation of grid voltage and current in the *dq*-frame, *P* and *Q* can be calculated as follows:

$$P = \frac{3}{2}(V_{gd} \cdot i_{gd} + V_{gq} \cdot i_{gq}) \tag{4}$$

$$Q = \frac{3}{2}(V_{gq} \cdot i_{gd} - V_{gd} \cdot i_{gq}) \tag{5}$$

The computed real and reactive powers comprise DC and oscillating components, and therefore, they are passed through the LPF to get the average component only.

Two PI regulators are employed to track the reference active (P_{cmd}) and reactive (Q_{cmd}) power commands. These PI regulators generate the reference active current component (i^*_{gd}) and reactive current component (i^*_{gq}), as follows:

$$i^*_{gd} = \left(K_{p1} + \frac{K_{i1}}{s}\right)(P_{cmd} - P) \tag{6}$$

$$i^*_{gq} = \left(K_{p2} + \frac{K_{i2}}{s}\right)(Q_{cmd} - Q) \tag{7}$$

where K_{p1}, K_{p2} and K_{i1}, K_{i2} are PI constants controller for control over real and reactive powers, respectively. P_{cmd} is the charging–discharging reference active power for battery and Q_{cmd} is the reference reactive power requested by the grid.

Furthermore, inverse transformation generates the periodic current reference current in a–b–c domain

$$\begin{bmatrix} i^*_a \\ i^*_b \\ i^*_c \end{bmatrix} = T^T \begin{bmatrix} i^*_{gd} \\ i^*_{gq} \end{bmatrix} \tag{8}$$

where

$$T^T = \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ \cos\left(\omega t - \frac{2\pi}{3}\right) & \sin\left(\omega t - \frac{2\pi}{3}\right) \\ \cos\left(\omega t - \frac{4\pi}{3}\right) & \sin\left(\omega t - \frac{4\pi}{3}\right) \end{bmatrix} \tag{9}$$

These reference currents (i^*_a, i^*_b and i^*_c) are regulated by using PR controllers. The error between the reference current and actual grid current is applied to the PR controller. The output signal of PR controller is superimposed on pulse-width modulator (PWM) carrier waveform to obtain the duty cycle (d_a, d_b and d_c) for the converter

$$d_{abc} = K_{pabc}(i^*_{abc} - i_{abc}) + K_{Rabc}\frac{2\omega_c S}{S^2 + 2\omega_c S + \omega^2}(i^*_{abc} - i_{abc}) \tag{10}$$

where K_{pa}, K_{pb} and K_{pc} are the proportional constants of PR controllers for phases a, b and c, respectively. Similarly, K_{Ra}, K_{Rb} and K_{Rc} are resonant constants and ω_c is the cut-off frequency. The design of PR current controller is discussed in the next section.

4 Current controller design

The primary objective of the proposed work is to evolve a versatile current control for EV charger. For better understanding, the proposed system is presented in Fig. 3. Here, R_f, L_f and C_f are filtering resistance, inductance and capacitance, respectively. The grid may act as a source or load depending on charger operating mode (i.e. G2V or V2G); accordingly, the grid inductance is represented by L_l and resistance is R_l.

The equivalent inner current control for the proposed controller is depicted in Fig. 4. Since the reference current signal is periodic in nature and accordingly a modified PR controller is employed to track the reference current signal.

The transfer function of the ideal PR controller in continuous time domain is given in the equation below:

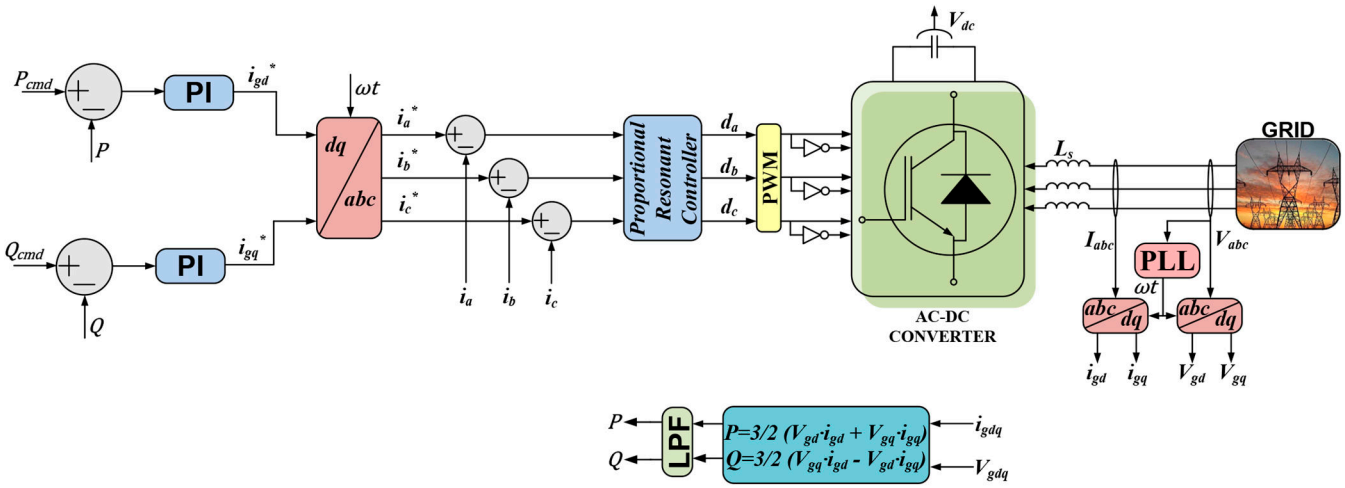


Fig. 2 *Control description of the off-board charger*

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

$$G_c(S) = K_p + \frac{2K_R S}{S^2 + \omega^2} \tag{11}$$

where K_p and K_R are the gains of PR controller with ω being the angular frequency. The transfer function of PWM in continuous time domain is given in the equation below:

$$G_{PWM}(S) = \frac{K_{PWM}}{1.5T_s S + 1} \tag{12}$$

where K_{PWM} is the corresponding PWM gain of three-phase converter and T_s is the switching time period.

Fig. 5 shows the frequency response of ideal PR controller $G_c(S)$ when $K_p = 1$, $K_R = 1$ and $\omega_c = 314$ rad/s. From Fig. 5, it can be observed that the gain value of ideal PR controller is infinite at fundamental frequency (ω).

The transfer function between the error signal (E) and reference current signal (I_s^*) of inner current closed-loop control system is given by

$$\frac{E(S)}{I_s^*(S)} = \frac{1}{1 + G(S)H(S)} \tag{13}$$

where $H(S) = 1$ as given closed-loop system is a unity feedback system and $G(S) = G_c(S) \cdot G_{PWM}(S) \cdot G_{PLANT}(S)$

$$\frac{E(S)}{I_s^*(S)} = \frac{X(S)}{K_{PWM}\left(K_p S^2 + 2K_R S + K_p \omega^2\right) + X(S)} \tag{14}$$

where $X(S) = (S^2 + \omega^2)(1.5T_s S + 1)(L_f + R_f)$.

If the reference input signal is $A\omega/(S^2 + \omega^2)$, then the error signal is

$$E(S) = \frac{A\omega(1.5T_s S + 1)(L_f + R_f)}{K_{PWM}\left(K_p S^2 + 2K_R S + K_p \omega^2\right) + X(S)} \tag{15}$$

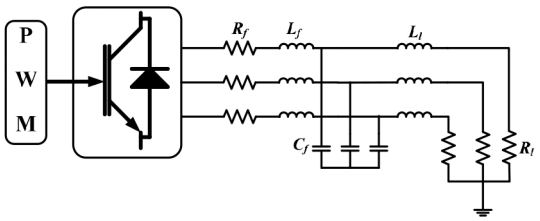


Fig. 3 *Proposed system*

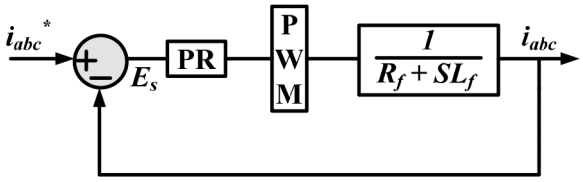


Fig. 4 *Inner current control loop*

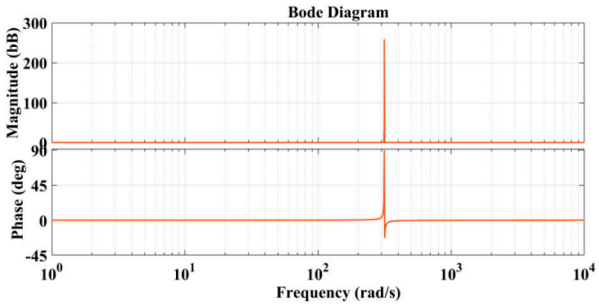


Fig. 5 *Bode plot of ideal PR controller*

According to the final value theorem, the steady-state value of the error signal is

$$\lim\ e(t) = \lim\ S\,E(S) = 0 \tag{16}$$

In ideal PR controller, the steady-state value of error is zero, but the gain at the resonance frequency is infinite. Practically, it cannot be possible to attain infinite gain at resonance frequency. Moreover, the infinite gain will fall quickly and may compromise the system stability. Owing to the quick fall, it is difficult to track the fluctuations in frequency and voltage of grid [26]. Therefore, for cases, modified PR controller would be utilised, whose transfer function is given in (18)

$$G_{mc}(S) = K_p + \frac{2K_R(\omega_c S + \omega_c^2)}{S^2 + 2\omega_c S + (\omega_c S + \omega_c^2)} \tag{17}$$

$$G_{mc}(S) \simeq K_p + \frac{2K_R \omega_c S}{S^2 + 2\omega_c S + \omega^2} \tag{18}$$

The transfer function of a modified PR controller has a frequency parameter (ω_c) in the denominator, which is responsible for modifying the magnitude and bandwidth at the fundamental frequency. The proportional gain factor K_p is responsible for determining the value of gain margin, phase margin and transient performance of the controller, whereas resonant gain factor K_R adjusts the controller's gain at the resonance frequency. The resonant bandwidth is determined by the cut-off frequency (ω_c). In general, the value of ω_c is smaller, i.e. between 5 and 15 rad/s [31]. The frequency response of a practical PR controller is presented in Fig. 6.

5 Simulation results

MATLAB/Simulink environment is utilised to simulate a 12.5 kVA off-board EV charger. To certify the transient and steady-state response of the proposed charger controller, a simulation scenario has been created in different modes as listed in Table 3. Each mode is simulated for 3 s. A 600 V lithium (Li)-ion battery with 50% SOC and a three-phase IGBT with antiparallel diode-based bidirectional AC–DC converter has been utilised in simulation. For performing the simulation system, parameters taken are listed in Table 1.

The simulation results of grid-side current (I_a, I_b and I_c), battery voltage (V_{bat}), battery current (I_{bat}) and measured active (P) and reactive (Q) powers during all operating modes are shown in Fig. 7.

During all operating modes, the battery voltage can be observed, which is almost constant. The battery current is positive during discharging operation and negative in case of charging. Since the apparent power of the charger is constant throughout all operating modes; therefore, the magnitude of the grid-side current is constant.

The zoomed version of changeovers from one operating mode to another is presented in Fig. 8. In the zoomed version, the three-phase grid current and phase a voltage (V_a) along with measured active and reactive powers at grid side is shown. First two modes are associated with charging and discharging operations only. During mode 1, the battery is charging from the grid and phase difference between grid voltage and current of phase a can be observed, which is exactly in phase. Similarly, phase a current is exactly out of phase with a voltage of phase a in case of discharging (i.e. mode 2). The transition between modes 1 and 2 is shown in Fig. 8*a*. During modes 3 and 4, only reactive power operation has been performed. During this, the current is exactly 90° lag behind the voltage and exactly lead by 90° in modes 3 and 4, respectively, and transition between modes 3 and 4 is shown in Fig. 8*c*. Modes 5–8 evolve the active and reactive power operations simultaneously. Modes 5 and 6 show the operation of reactive power while charging the battery. During mode 5, the active power transfer to the battery is 10 kW and inductive reactive power

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

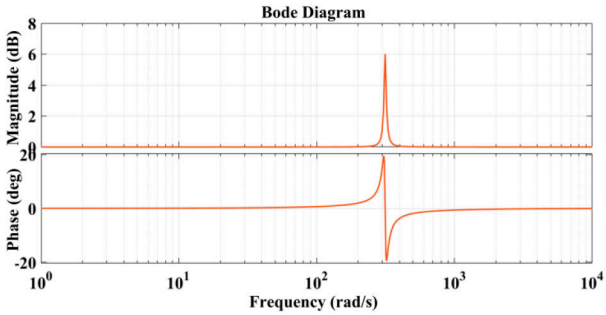


Fig. 6 *Bode plot of practical PR controller*

Table 3 Simulation scenario

Mode	Active power, kW	Reactive power, kVAR	Apparent power, kVA	Power factor	Time, s
1	12.5	0	12.5	unity	0–3
2	−12.5	0	12.5	−1	3–6
3	0	12.5	12.5	0	6–9
4	0	−12.5	12.5	0	9–12
5	10	7.5	12.5	0.8 (lag)	12–15
6	7.5	−10	12.5	0.6 (lead)	15–18
7	−5.5	11.2	12.5	0.44 (lag)	18–21
8	−11.2	−5.5	12.5	0.89 (lead)	21–24

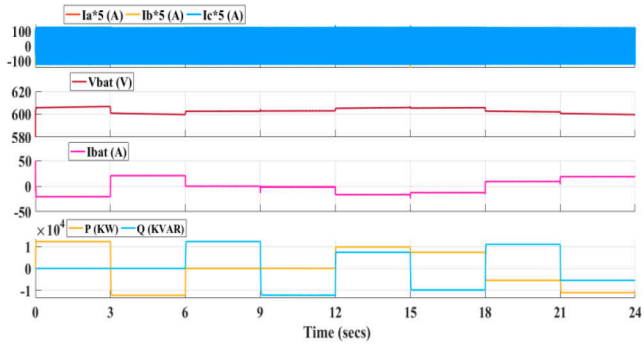


Fig. 7 *Measured grid-side current, battery voltage, battery current and active–reactive power*

compensation is 7.5 kVA. During this mode, the current is lagging behind by around 36° with the voltage. Similarly, the phase angle between voltage and current during mode 6 is around 53° lead and the transition between modes 5 and 6 is shown in Fig. 9*a*. The last two modes are associated with reactive power compensation during battery-discharging operation. During mode 7, active power supply to the grid is 5.5 kW and inductive reactive power compensation is 11.2 kVA. The angle between voltage and current during this mode is around 116° lag, and similarly, lead with around 206° during mode 8.

The controller responsible for both transient and steady is shown under simulation. It has been validated that the controller performs tremendously during all possible operating modes of an EV charger. The transient response of the controller is also good and takes less than three grid cycles to settle down. Furthermore, this controller has been implemented for scaled-down hardware prototype in real time.

6 Experimental results and discussion

Fig. 10*a* shows the experimental setup and Fig. 10*b* shows the battery bank used for charging–discharging purpose. To validate the controller's performance in real time, a scaled-down hardware prototype has been developed. For this purpose, three IGBT legs (each having two IGBTs with antiparallel diode) AC–DC bidirectional converter has been designed. A series of eight 12 V Li-ion battery are connected across the DC-link capacitor of

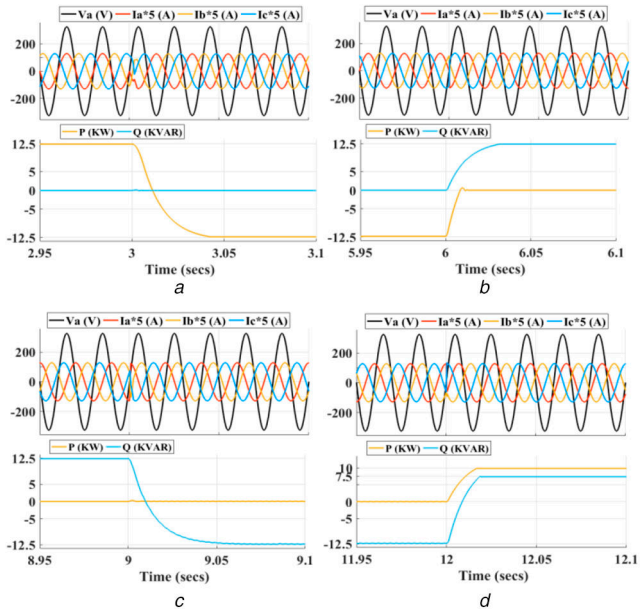


Fig. 8 *Measured three-phase grid current and phase a grid voltage during the transition from*
(a) Mode 1 to 2, ***(b)*** Mode 2 to 3, ***(c)*** Mode 3 to 4, ***(d)*** Mode 4 to 5

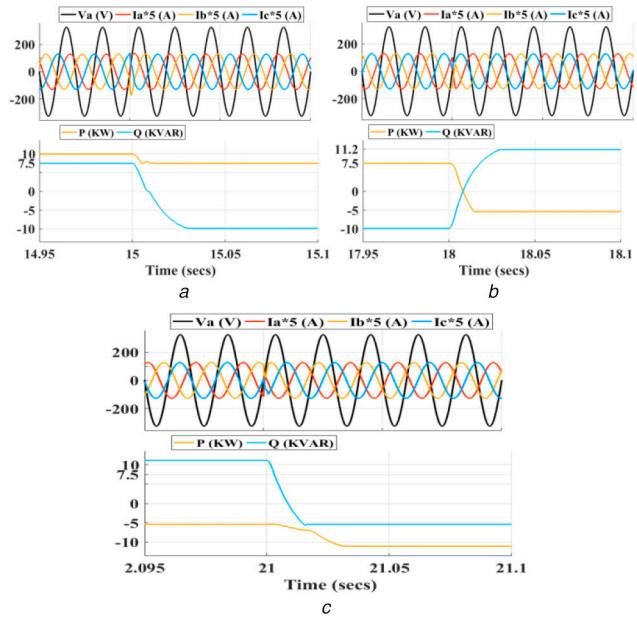


Fig. 9 *Measured three-phase grid current and phase a grid voltage during the transition from*
(a) Mode 5 to 6, ***(b)*** Mode 6 to 7, ***(c)*** Mode 7 to 8

converter. The experimental parameters are listed in Table 1. A *dSPACE 1104* is utilised to carry out the experimental study.

The charger's performance has been tested as per the scenario listed in Table 4 and outcomes are presented in Fig. 9.

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

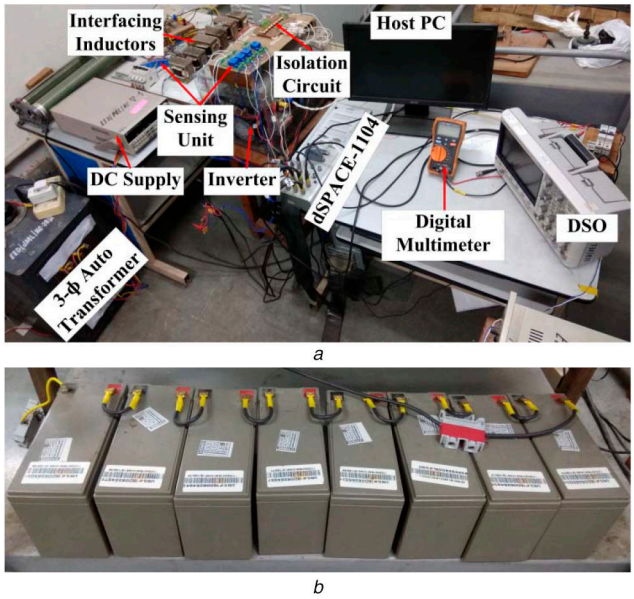


Fig. 10 *Hardware Prototype*
(a) Setup, ***(b)*** Battery Bank

Table 4 Experimental scenario

Mode	Active power, W	Reactive power, VAR	Apparent power, VA	Power factor
1	350	0	350	unity
2	−350	0	350	−1
3	0	350	350	0
4	0	−350	350	0
5	280	210	350	0.8 (lag)
6	210	−280	350	0.6 (lead)
7	−150	316	350	0.42 (lag)
8	−316	−150	350	0.9 (lead)

The transition of three-phase grid current (I_{abc}), phase a voltage (V_a), powers (P and Q), battery voltage (V_{bat}) and battery current (I_{bat}) are presented in results. For safety purposes, the maximum apparent power of the charger is limited to 350 VA, while during simultaneous operation, the maximum amount of active power transfer is 316 W whereas reactive power compensation is 316 VAR.

Similar to the simulation scenario, the first two modes of experimental are associated with active power transfer without reactive power compensation. While transferring 350 W to battery from grid in mode 1, accordingly, there is no phase difference between grid voltage and current. Similarly, the charger performs V2G operation in mode 2 and transfer 350 W to the grid from battery. The phase difference between grid current and voltage is exactly 180°, and transitions of grid voltage of phase a (V_a) and three-phase grid current (I_{abc}) are shown in Fig. 11*a* and changeovers of P, Q, V_{bat}, I_{bat} are represented in Fig. 11*b*. Modes 3 and 4 are associated with 350 VAR of reactive power compensation (i.e. inductive and capacitive) without transferring of active power. During these operations, the battery current is constant and nearly equal to zero. The phase difference is exactly 90° in both the operating modes, but lags in mode 3 and leads in mode 4. Fig. 11*c* demonstrates the changeover of the current and voltage phase angle and Fig. 11*d* demonstrates the transitions of P, Q, V_{bat}, I_{bat} from mode 3 to 4.

Modes 5 and 6 perform both active and reactive power operations simultaneously. During mode 5, the amount of active power supply to the EV battery is 280 W and compensation of inductive reactive power is 210 VAR. During this mode, the grid current lags behind 36° by grid voltage. In mode 6, active and reactive powers are 210 W and −280 VAR, respectively. The grid current lead by 53° and transition from mode 5 to 6 is shown in Figs. 12*a* and *b*. Last two modes are associated with reactive power compensation during discharging operation of battery. During mode 7, battery supply 150 W power while compensating 316 VAR inductive reactive power, where 316 W supplying while compensating 150 VAR capacitive reactive power in mode 8. The phase angles during modes 7 and 8 are 115° and 205°, respectively, and transition from mode 7 to 8 is shown in Figs. 12*c* and *d*.

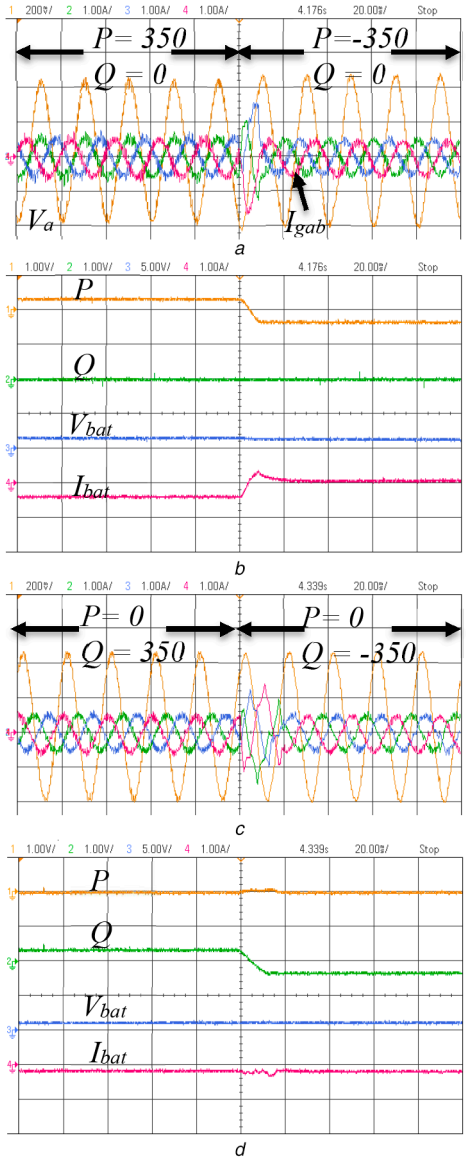


Fig. 11 *Experimental results transition from*
(a), (b) Mode 1 to 2, ***(c), (d)*** Mode 3 to 4

The hardware results show the transient and steady-state dynamics of the proposed controller. It can be seen that the controller has a very fast dynamic response and tracks the reference command in less than three grid cycles in real time also.

7 Conclusion

The control design of a single-stage three-phase off-board charger is simulated and is verified under real time. The performance of the EV charger is evaluated with its ability to work under all the four quadrants of the P–Q-plane. The proposed controller has two PI controllers for regulating active–reactive powers in the outer loop while the controlling of the grid current is being done via the PR controller in the inner control loop. The performance of the proposed controller for EV charger has been tested and validated in all eight different modes with an abrupt change in reference power. It has been observed that the controller responds quickly and takes less than three grid cycles to settle down. The performance has been validated through the MATLAB simulation and hardware

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

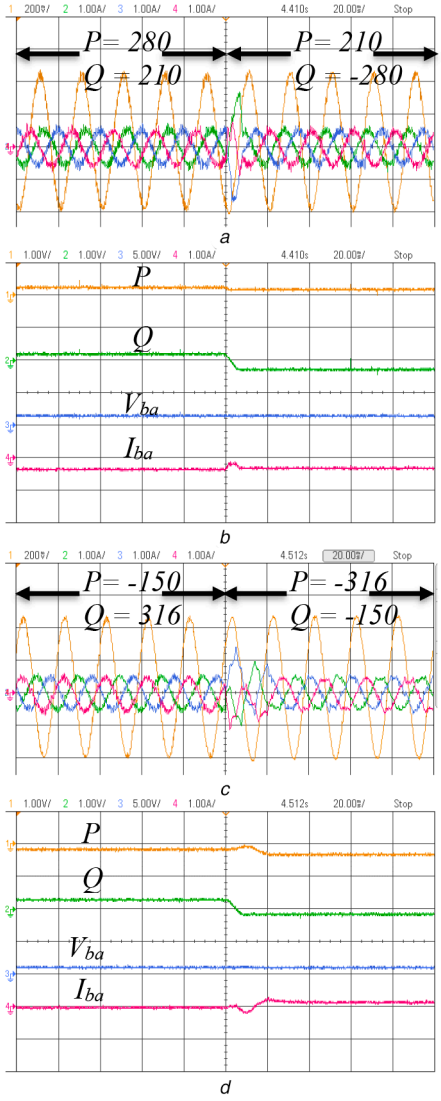


Fig. 12 *Experimental results transition from* ***(a), (b)*** Mode 5 to 6, ***(c), (d)*** Mode 7 to 8

prototype. The outcome of the proposed work justifies the efficient behaviour of the controller with experimental validation.

8 References

[1] Khalid, M.R., Alam, M.S., Sarwar, A., *et al.*: 'A comprehensive review on electric vehicles charging infrastructures and their impacts on power quality of the utility grid', *eTrans*, 2019, **1**, (100006), pp. 1–10

[2] Khan, W., Ahmad, F., Alam, M.S.: 'Fast EV charging station integration with grid ensuring optimal and quality power exchange', *Eng. Sci. Technol. Int. J.*, 2019, **22**, (1), pp. 143–152

[3] Pistoia, G.: '*Electric and hybrid vehicles*' (Elsevier, Italy, 2010)

[4] Chan, C.C., Bouscayrol, A., Chen, K.: 'Electric, hybrid, and fuel-cell vehicles: architectures and modeling', *IEEE Trans. Veh. Technol.*, 2010, **59**, (52), pp. 589–598

[5] González, L.G., Siavichay, E., Espinoza, J.L.: 'Impact of EV fast-charging stations on the power distribution network of a Latin American intermediate city', *Renew. Sustain. Energy Rev.*, 2019, **107**, pp. 309–318

[6] Williamson, S.S., Rathore, A.K., Musavi, F.: 'Industrial electronics for electric transportation: current state-of-the-art and future challenges', *IEEE Trans. Ind. Electron.*, 2015, **62**, (5), pp. 3021–3032

[7] Kesler, M., Kisacikoglu, M.C., Tolbert, L.M.: 'Vehicle-to-grid reactive power operation using plug-in electric vehicle bidirectional off-board charger', *IEEE Trans. Ind. Electron.*, 2014, **61**, (12), pp. 6778–6784

[8] Yilmaz, M., Krein, P.T.: 'Review of battery charger topologies, charging power levels and infrastructure for plug-in electric and hybrid vehicles', *IEEE Trans. Power Electron.*, 2013, **28**, (5), pp. 2151–2169

[9] Lu, X., Iyer, K.L.V., Lai, C., *et al.*: 'Design and testing of a multi-port sustainable DC fast-charging system for electric vehicles', *Electr. Power Compon. Syst.*, 2016, **44**, (14), pp. 1576–1587

[10] Hartmann, M., Friedli, T., Kolar, J.W.: 'Three-phase unity power factor mains interfaces of high-power EV battery charging systems'. Proc. ECPE Workshop Power Electronic Charging Electric Vehicle, Valencia, Spain, 21–22 March 2011, pp. 1–66

[11] Revolutionizing fast charging for electric vehicles, Intel, Santa Clara, CA, USA, July 2013

[12] Musavi, F., Edington, M., Eberle, W., *et al.*: 'Evaluation and efficiency comparison of front end AC–DC plug-in hybrid charger topologies', *IEEE Trans. Smart Grid*, 2012, **3**, (1), pp. 413–421

[13] Kisacikoglu, M.C., Ozpineci, B., Tolbert, L.M.: 'EV/PHEV bidirectional charger assessment for V2G reactive power operation', *IEEE Trans. Power Electron.*, 2013, **28**, (12), pp. 5717–5727

[14] Tan, K.M., Padmanaban, S.K., Yong, J.Y., *et al.*: 'A multi-control vehicle-to-grid charger with bi-directional active and reactive power capabilities for power grid support', *Energy*, 2019, **171**, (15), pp. 1150–1163

[15] Kisacikoglu, M.C., Kesler, M., Tolbert, L.M.: 'Single-phase on-board bidirectional PEV charger for V2G reactive power operation', *IEEE Trans. Smart Grid*, 2015, **6**, (2), pp. 767–775

[16] Kisacikoglu, M.C., Ozpineci, B., Tolbert, L.M.: 'Examination of a PHEV bidirectional charger for V2G reactive power compensation'. Proc. IEEE Applied Power Electronics Conf. Exposition (APEC), Palm Springs, CA, USA, 21–25 February 2010, pp. 458–465

[17] Wang, Y., John, T., Xiong, B.: 'A two-level coordinated voltage control scheme of electric vehicle chargers in low-voltage distribution networks', *Electr. Power Syst. Res.*, 2019, **168**, pp. 218–227

[18] Li, Y., Li, L., Peng, C., *et al.*: 'An MPC-based optimized control approach for EV-based voltage regulation in distribution grid', *Electr. Power Syst. Res.*, 2019, **172**, pp. 152–160

[19] Verma, A., Singh, B.: 'Three-phase off-board bi-directional charger for EV with V2G functionality'. 2017 Seventh Int. Conf. Power Systems (ICPS), Pune, India, 2017, pp. 145–150

[20] Singh, M., Khadkikar, V., Chandra, A., *et al.*: 'Grid interconnection of renewable energy sources at the distribution level with power-quality improvement features', *IEEE Trans. Power Deliv.*, 2011, **26**, (1), pp. 307–315

[21] Pandove, G., Trivedi, A., Singh, M.: 'Repetitive control-based single-phase bidirectional rectifier with enhanced performance', *IET Trans. Power Electron.*, 2016, **9**, (5), pp. 1029–1036

[22] Cortés, P., Rodríguez, J., Antoniewicz, P., *et al.*: 'Direct power control of an AFE using predictive control', *IEEE Trans. Power Electron.*, 2008, **23**, (5), pp. 2516–2523

[23] He, T., Zhu, J., Lu, D.D., *et al.*: 'Modified model predictive control for bidirectional four-quadrant EV chargers with extended set of voltage vectors', *IEEE J. Emerg. Sel. Top. Power Electron.*, 2019, **7**, (1), pp. 274–281

[24] Lu, X., Iyer, K.L.V., Mukherjee, K., *et al.*: 'A dual-purpose triangular neural network-based module for monitoring and protection in bi-directional off-board level-3 charging of EV/PHEV', *IEEE Trans. Smart Grid*, 2012, **3**, (4), pp. 1670–1678

[25] Haddoun, A., Benbouzid, M.E.H., Diallo, D., *et al.*: 'Modeling, analysis, and neural network control of an EV electrical differential', *IEEE Trans. Ind. Electron.*, 2008, **55**, (6), pp. 2286–2294

[26] Teodorescu, R., Blaabjerg, F., Liserre, M., *et al.*: 'Proportional resonant controllers and filters for grid-connected voltage-source converters', *IEE Proc. Electr. Power Appl.*, 2006, **153**, (5), pp. 750–762

[27] Jiao, R., Zhu, J., Ma, L., *et al.*: 'Deadbeat instantaneous power control for off-board charger of electric vehicle for V2G application', *Int. J. Hydrog. Energy*, 2017, **42**, (29), pp. 18174–18180

[28] Lucas, A., Bonavitacola, F., Kotsakis, E., *et al.*: 'Grid harmonic impact of multiple electric vehicle fast charging', *Electr. Power Syst. Res.*, 2015, **127**, pp. 13–21

[29] Kesler, M., Ozdemir, E.: 'Synchronous reference frame (SRF) based control method for UPQC under unbalanced and distorted load conditions', *IEEE Trans. Ind. Electron.*, 2011, **58**, (9), pp. 3967–3975

[30] Yazdani, A., Iravani, R.: '*Voltage-sourced converters in power systems: modeling, control, and applications*' (Wiley-IEEE Press, USA, 2010)

[31] Bo, H.: '*Research on the single-phase PWM rectifier*' (Huazhong University of Science & Technology, People's Republic of China, 2012)

© The Institution of Engineering and Technology 2020

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:40:19 UTC from IEEE Xplore. Restrictions apply.

International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 3, March 2020, pp. 371-380, Article ID: IJARET_11_03_032
Available online athttp://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=3
ISSN Print: 0976-6480 and ISSN Online: 0976-6499

© **IAEME** Publication **Scopus** Indexed

RESPONSE ANALYSIS OF PLUS SHAPED TALL BUILDING WITH DIFFERENT BRACING SYSTEMS UNDER WIND LOAD

Dr. Ritu Raj
Assistant Professor, Department of Civil Engineering
Delhi Technological University

Shubhangi Jha, Shreyansh Singh and Siddhant Choudhary
Bachelor of Technology, Department of Civil Engineering
Delhi Technological University

ABSTRACT

This paper aims at response study and analysis of plus shaped tall building with different orientations of bracing systems under wind loads. The effect of wind load on building becomes very critical with increase in height of tall buildings. As not much encouraging information is available in the standard codes of practice regarding tall buildings with irregular plans and cross-sectional shapes, hence, more research needs to be done in the given area. With the same objective, the present study focuses on a different plan i.e. plus shaped tall building exposed to 0°, 30° and 60° angles of attack of wind. Isolated condition (without bracing system), V bracing, Cross bracing, Single Diagonal bracing and Inverted V bracing system have been considered to analyse the effectiveness of various bracing systems as structural system against wind loads in tall buildings. Bentley STAAD Pro software v8i module was used to carry out response study. Prototype buildings are presumed to be constituted of RCC beams and columns. The prototype building was designed as G+35 with 4.5 m ground floor height and 3.3 m remaining floors' height. Mean response of prototype building at windward position and leeward position for 0°, 30° and 60° wind incidence angles including base shear (F_x), moment about y axis (M_y), twisting moment (M_z) and deflection in x direction has been obtained to study the outcome under wind loads of plus shape and different bracing systems.

Key words: Wind Loads, Tall Buildings, Plus shaped buildings, Cross Bracing, Single Diagonal Bracing, V Bracing, Inverted V Bracing

Cite this Article: Dr. Ritu Raj, Shubhangi Jha, Shreyansh Singh and Siddhant Choudhary, Response Analysis of Plus Shaped Tall Building with Different Bracing Systems Under Wind Load, *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(3), 2020, pp 371-380.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=3

1. INTRODUCTION

Tremendous rise in population over past few years has added undue stress on the limited land for accommodation resulting in shift from horizontal to vertical mode of expansion. As a result, tall buildings and their judicious design has become an imperative answer to the question of efficiently utilizing land for residential, industrial, recreational, educational and other purposes. During the design of tall buildings, the correct estimation of lateral loads especially wind loads becomes very crucial as wind is a complicated occurrence that varies randomly.

However, various standard codes of practices being used worldwide for estimation of wind loads although provide some information but are not exhaustive. [1-5] They deal with no shapes other than standard cross-sectional shapes including square shape and rectangular shape and give very minimal information on pressure distribution on tall buildings under wind loads at skew angles of attack.

Review of research work done in the field shows that so far majority of the focus has been on pressure distributions of the tall building models only. S. Chakraborty et al, 2014 used wind tunnel experiment to study mean wind pressure coefficients on an irregular plus shaped tall building at wind incidence angle 0° and 45° respectively. Yi Yi et al, 2017 derived a formula for estimation of wind induced torques on L shaped tall buildings. R. Sheng et al, 2018 studied effects of global and local wind loads on high-rise building through wind tunnel tests and concluded that wall pressure forces depend on the location. [6-9]

A.K. Mulla and B.N. Srinivas, 2015 executed response analysis of a tall R.C. structure with outrigger system and steel bracing using ETABS program under static and dynamic loads. [10] M. Boostani et al, 2018 contemplated an experimental program using fem (finite element method) numerical examination to propose supporting frameworks called 'o' grid bracing systems, for seismic tremor safe steel structures. [11] A. Arzeytoon and V. Toufigh, 2018 conducted probabilistic seismic performance assessment of ribbed bracing systems. [12] A. Rahimi and M.R. Maheri, 2018 considered the impacts of retrofitting rc frames by x-bracing on the performance of columns under earthquake loads. [13] Since, most of the work related to bracings focused on seismic load analysis; hence, a need to study the response of tall buildings with different types of concentric bracing systems under wind loads was realized.

2. METHODOLOGY

2.1 Response study technique

The present study is focused on wind load response analysis of a plus shaped tall building. Data from boundary layer wind tunnel testing experiments was used to calculate forces acting on the models of uniform area along the height. The different cases considered for study included isolated condition (plus shaped building without any bracing system), plus shaped building with single diagonal bracing, V-bracing, cross bracing and inverted V-bracing respectively.

The columns under study of the building were named Column A (Column at windward position) and column B (column at leeward position) as shown in Figure 1.

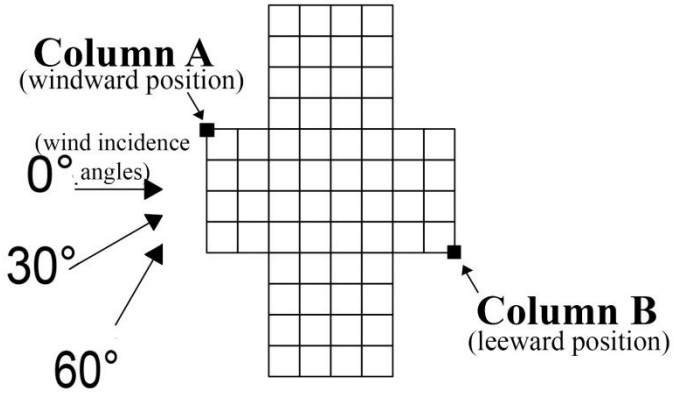


Figure 1. Windward (Column A) and Leeward (Column B) positions in the plus shaped building.

Prototype buildings were subjected to 0°, 30° and 60° angles of attack and parameters such as Base Shear (F_x), Moment in the y-direction (M_y), Twisting moment (M_z) and Deflection in the x-direction (X) was studied at windward side and leeward side, respectively. Readily available software package STAAD Pro v8i was used for analysis.

2.2. Details of the Prototype Building

- Floor Area = 1600 m^2
- Height = 120 m
- Number of floors = 36 (G+35)

3. RESULTS AND DISCUSSION

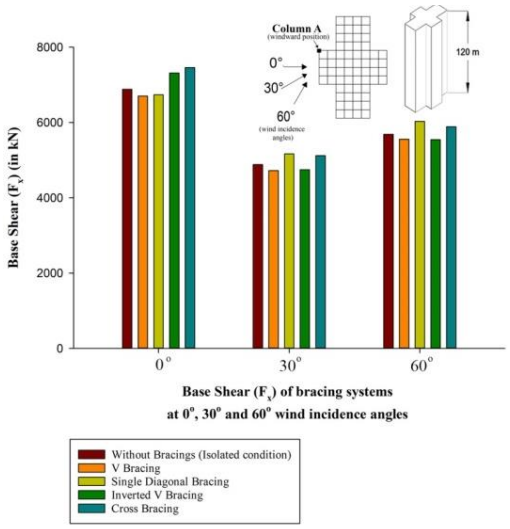


Figure 2. Impact of different wind attack angles on F_x at Windward Side (column A) in a Plus shaped Building

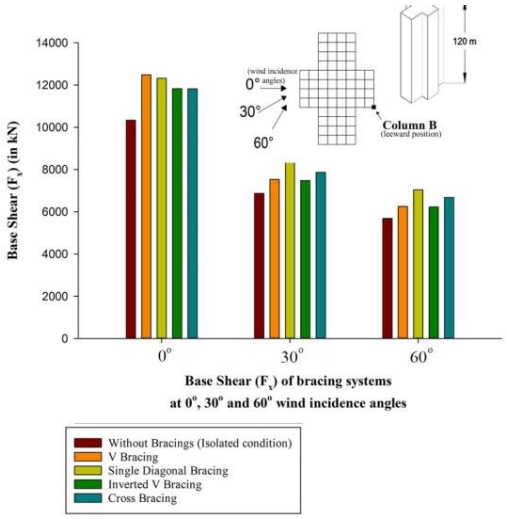


Figure 3. Impact of different wind attack angles on F_x at Leeward Side (column B) in a Plus shaped Building

Base shear at 0° was higher than 30° and 60° angles of attack in both cases i.e. windward side and leeward side of the building. However, at 0° wind incidence angle, cross bracing exhibited maximum base shear at windward side (7462.84 kN) while V bracing exhibited maximum base shear at leeward side (12480.57 kN). At 30° and 60° wind incidence angles, single diagonal bracing system exhibited maximum value of base shear at both windward and leeward positions of the building.

For windward position of the building (column A), V bracing system showed minimum values of base shear at all angles of attack with least value of base shear being 4722.85 kN at 30° angle of attack, whereas for leeward position of the building (column B), isolated condition (without any bracing system) showed minimum values of base shear at all angles of attack with least value of base shear as 5685.17 kN at 60° angle of attack.

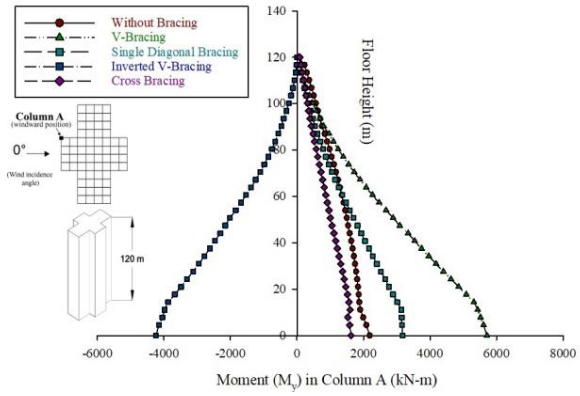


Figure 4. Effect of 0° incidence angle on moment (M_y) at Windward Side (column A) in a Plus shaped Building

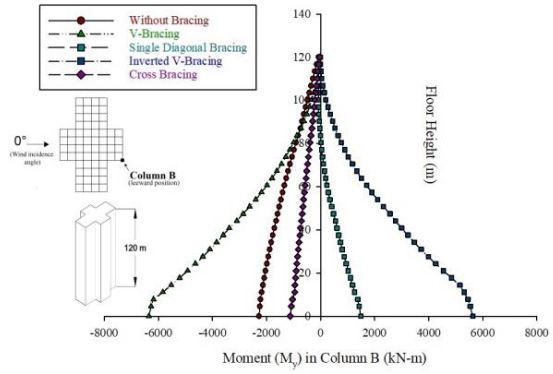


Figure 5. Effect of 0° incidence angle on moment (M_y) at Leeward Side (column B) in a Plus shaped Building

Base moment was seen to be maximum at 0-degree angle of wind attack for both column-A (windward position) and Column B (Leeward position), as observed in Figures 4 and 5. It was gauged that base moment (M_y) in an isolated condition was maximum at 60 degree angle of attack for column-A (windward position) and at 30 degree angle of attack for Column B (Leeward position).

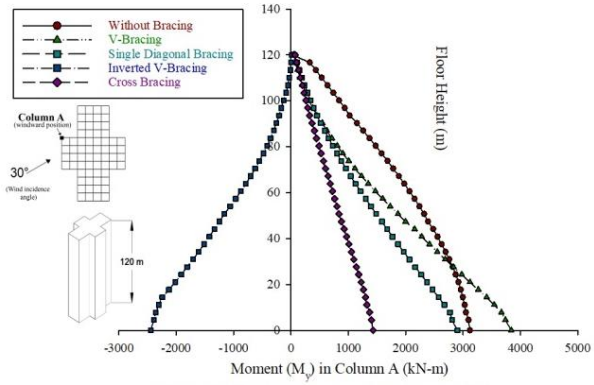


Figure 6. Effect of 30° incidence angle on moment (M_y) at Windward Side (column A) in a Plus shaped Building

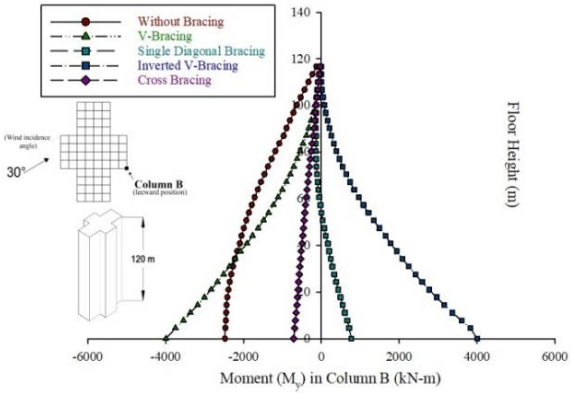


Figure 7. Effect of 30° incidence angle on moment (M_y) at Leeward Side (column B) in a Plus shaped Building

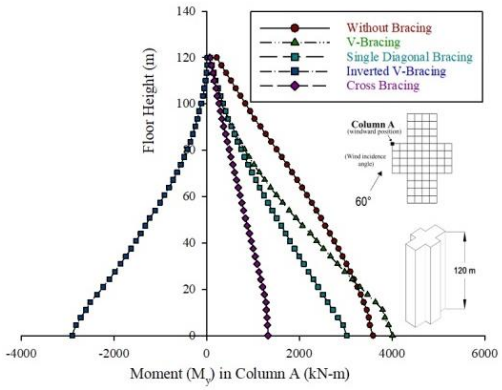


Figure 8. Effect of 60° incidence angle on moment (M_y) at Windward Side (column A) in a Plus shaped Building

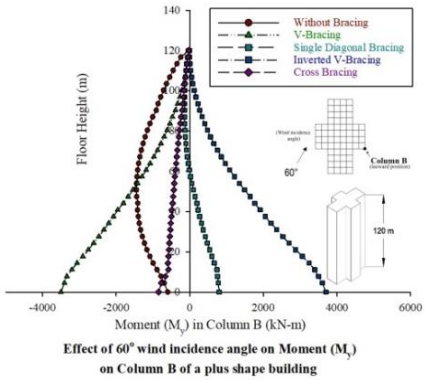


Figure 9. Effect of 60° incidence angle on moment (M_y) at Leeward Side (column B) in a Plus shaped Building

At all angles of attack, Inverted V bracing showed minimum values of M_y at windward position and maximum values of M_y at leeward position while V bracing showed maximum values of M_y at windward position and minimum values of M_y at leeward position.

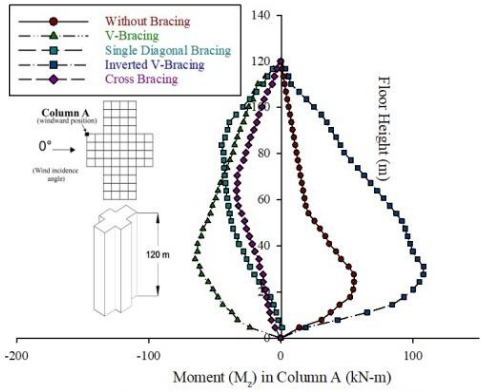


Figure 10. Effect of 0° incidence angle on moment (M_z) at Windward Side (column A) in a Plus shaped Building

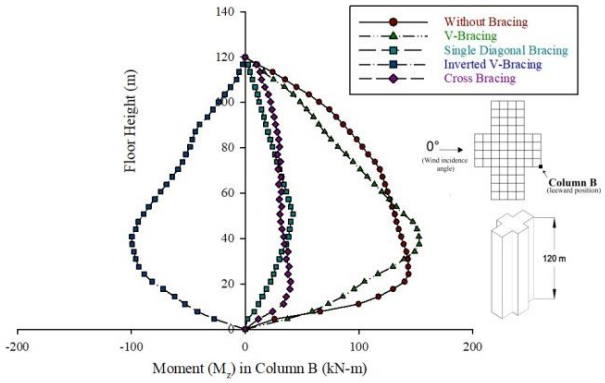


Figure 11. Effect of 0° incidence angle on moment (M_z) at Leeward Side (column B) in a Plus shaped Building

For windward position of the building (Column A), Inverted V bracing system showed maximum values of twisting moment at all angles of attack (with maximum value being 108.20 kN-m at 0° wind incidence angle and 25% building height), while V bracing system showed minimum values of twisting moment at all angles of attack (with least value being -62.32 kN-m at 30° wind incidence angle and 33.34% building height).

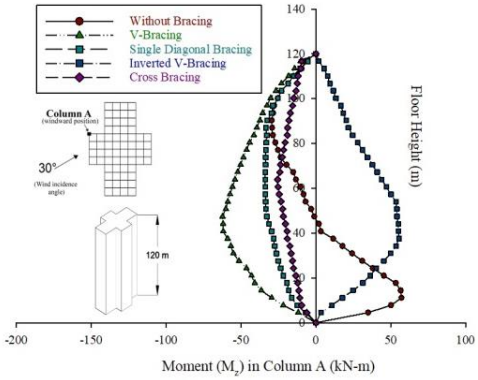


Figure 12. Effect of 30° incidence angle on moment (M_z) at Windward Side (column A) in a Plus shaped Building

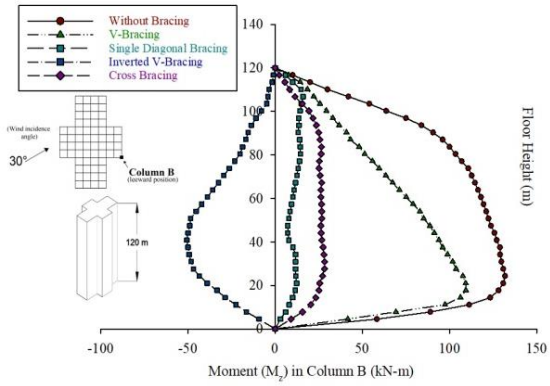


Figure 13. Effect of 30° incidence angle on moment (M_z) at Leeward Side (column B) in a Plus shaped Building

For leeward position (Column B), isolated condition (without any bracing system) and V bracing system showed similarly high values of twisting moment at all angles of attack (with

maximum value of V bracing system as 152.80 kN-m at 0° wind incidence angle and 33.34% building height), while Inverted V bracing system showed minimum values of twisting moment at all angles of attack (with least value being -99.67 kN-m at 0° wind incidence angle and 33.34% building height).

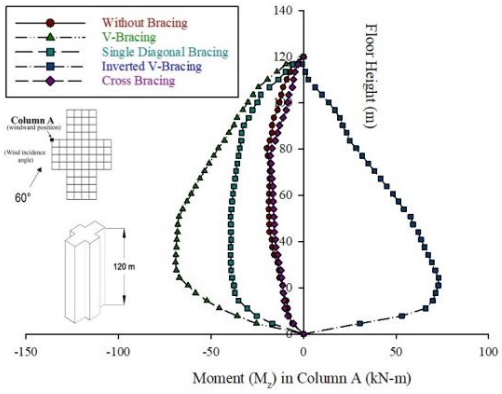


Figure 14. Effect of 60° incidence angle on moment (M_z) at Windward Side (column A) in a Plus shaped Building

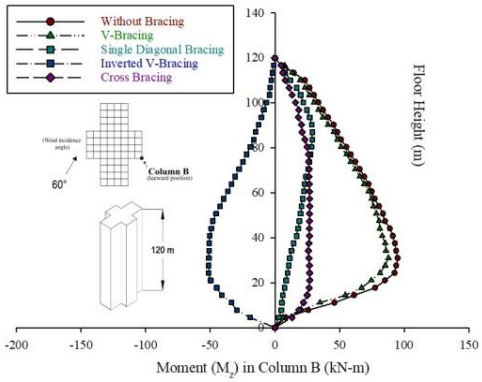


Figure 15. Effect of 60° incidence angle on moment (M_z) at Leeward Side (column B) in a Plus shaped Building

Numerical values obtained for twisting moment were negligible in all conditions as compared to corresponding values of base moment about Y-axis. It was hence concluded that a section designed for maximum axial force or base moment is safe and can take care of the twisting moment. Therefore, there is no need to design the section of a column separately for twisting moment.

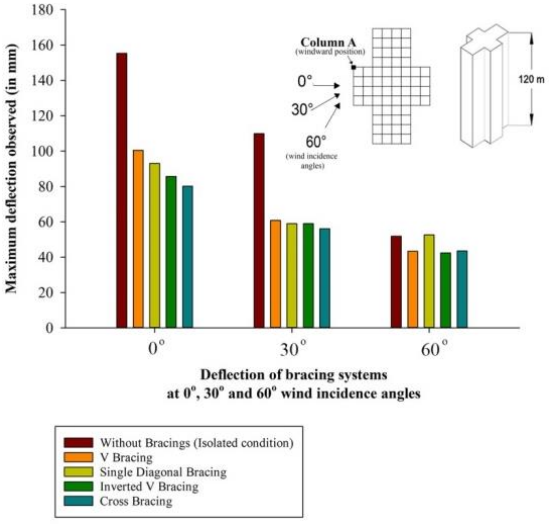


Figure 16. Impact of different wind attack angles on Deflection at Windward Side (column A) in a Plus shaped Building

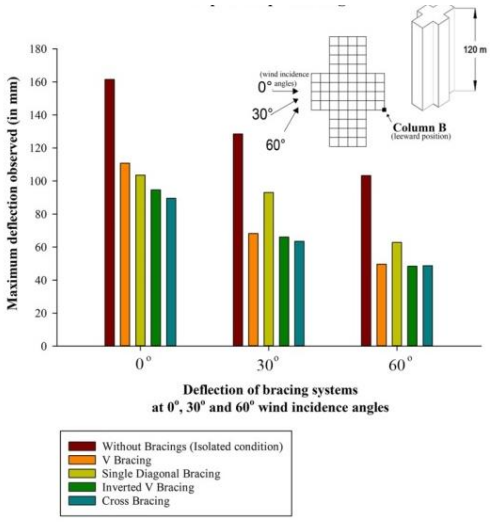


Figure 17. Impact of different wind attack angles on Deflection at Leeward Side (column B) in a Plus shaped Building

Deflection in isolated condition was seen to be maximum at all angles of attack with highest value of deflection as 161.46 mm at leeward position and 0° wind incidence angle. For both i.e. leeward and windward positions of the building, inverted V bracing had the minimum value of deflection at all angles of attack with least value of 42.35 mm at 60° wind incidence angle at windward position.

At windward position, reduction of sway by 44.8%, 46.33% and 18.29% while at leeward position, reduction of sway by 41.37%, 48.61% and 53.17% at 0°, 30° and 60° wind incidence angles respectively was observed due to Inverted V bracing.

5. CONCLUSIONS

1. Single diagonal bracing system reflected lesser axial force values. The axial force values show a very slow decrease from bottom to 30% height of the building and then a rapid decrease to the top.

2. Twisting moment was observed to be negligible in all systems, except in the case of inverted V-bracing system for column B (Leeward position) at 60-degree angle of attack.

3. It was concluded that, at 0-degree, 30 degree and 60-degree angle of wind attack, V and inverted V-bracing systems gave comparable values with minimum sway of 67%, 50% and 54% respectively.

REFERENCES

[1] AS/NZS: 1170.2 (2002), "Structural Design Actions, Part-2: Wind Action"

[2] ASCE: 7-02 (2002), "Minimum Design Loads for Buildings and Other Structures"

[3] BS: 63699 (1995), "Loading for Buildings: Part 2 – Code of Practice for Wind Loads"

[4] EN 1991-1-4 (2005), "Euro code 1: Actions on Structures - Wind Actions"

[5] IS:875-Part-3 (2015), "Code of Practice for Design Loads (other than Earthquake Loads) for Buildings and Structures- Wind Loads"

[6] Chakraborty, S., Dalui, S.K. and Ahuja, A.K., 2014. Wind load on irregular plan shaped tall building-a case study. Wind and Structures, 19(1), pp.59-73.

[7] Raj, R., Sharma, A. and Chauhan, S., 2018. Response of Square and Plus Shaped Buildings on Varying Wind Loads. Journal of Structural Engineering.

[8] Li, Y., Li, Q.S. and Chen, F., 2017. Wind tunnel study of wind-induced torques on L-shaped tall buildings. Journal of Wind Engineering and Industrial Aerodynamics, 167, pp.41-50.

[9] Sheng, R., Perret, L., Calmet, I., Demouge, F. and Guilhot, J., 2018. Wind tunnel study of wind effects on a high-rise building at a scale of 1: 300. Journal of Wind Engineering and Industrial Aerodynamics, 174, pp.391-403.

[10] Mulla, A.K. and Srinivas, B.N., 2015. A study on outrigger system in a tall RC structure with steel bracing. International Journal of Engineering Research and, 4.

[11] Boostani, M., Rezaifar, O. and Gholhaki, M., 2018. Introduction and seismic performance investigation of the proposed lateral bracing system called "OGrid". Archives of civil and mechanical engineering, 18(4), pp.1024-1041.

[12] Arzeytoon, A. and Toufigh, V., 2018. Probabilistic seismic performance assessment of ribbed bracing systems. Journal of Constructional Steel Research, 148, pp.326-335.

[13] Rahimi, A. and Maheri, M.R., 2018. The effects of retrofitting RC frames by X-bracing on the seismic performance of columns. Engineering Structures, 173, pp.813-830.

[14] IS:456 (2000), "Plain and Reinforced Concrete – Code of Practice"

[15] IS:875-Part-1 (1987), "Code of Practice for Design Loads (other than Earthquake Loads) for Buildings and Structures- Dead Loads"

[16] IS:875-Part-2 (1987), "Code of Practice for Design Loads (other than Earthquake Loads) for Buildings and Structures- Imposed Loads"

Environment, Development and Sustainability
https://doi.org/10.1007/s10668-020-00741-x

Selection of sustainable solutions for crop residue burning: an environmental issue in northwestern states of India

Pravin Kumar[1] **· Rajesh Kumar Singh**[2]

Received: 15 May 2019 / Accepted: 15 April 2020
© Springer Nature B.V. 2020

Abstract

The main purpose of the paper is to find the sustainable solutions for crop residues burning in the northwestern states of India. Pre- and post-monsoon burning of the crop residues in Punjab, Haryana, Rajasthan, and western Uttar Pradesh increases the particulate matter significantly in the environment. During this period, the wind remains stagnant in Delhi NCR (National Capital Region) which creates an airlock situation in this region. The air quality during this period becomes very poor. It is a major health issue in the northern part of the country, especially in NCR Delhi which is densely populated and industrialized. It has become a major concern for the state as well as central government. Therefore, the purpose of this study is to review the factors forcing the farmers to burn the crop residues in the field and analyze the solutions considering the social and economic sustainability of the farmers. A focus group discussion composed of farmers, academicians, and industry experts has been used to get the opinion regarding the solutions for crop residues burning in the field. For an analysis of the information collected, a hybrid approach of analytic hierarchy process and elimination and choice expressing reality have been used. Based on the analysis, it is observed that use of crop residues for the preparation of biogas, biochar, and bioethanol is the most preferred substitute, whereas use as fuel for the thermal power plant is the second preferred substitute for use of crop residues.

Keywords Greenhouse gases · Particulate matters · Rice and wheat crop residue-burning · Air pollution · Multi-criteria decision making · AHP · ELECTRE

✉ Pravin Kumar
pravin.dce@gmail.com

Rajesh Kumar Singh
rksdce@yahoo.com

[1] Department of Mechanical Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India

[2] Management Development Institute, Gurgaon, Haryana 122007, India

Published online: 23 April 2020

1 Introduction

Agriculture is a way for the livelihood of a large section of the people of India. Haryana and Punjab are the major producers of rice and wheat. These two crops are also produced in some part of the western Uttar Pradesh and Rajasthan. In India, the annual production of crop residue is approx. 516.73 million tonnes (Mt), out of which 110, 122, 71, 26, 141, 8, and 28 Mt of are produced from production of wheat, rice, maize, millets, sugarcane, fiber crops, and pulses, respectively (Devi et al. 2017, MOSPI, 2013–2014). A major part of the surplus crop residue is burnt in the field. It was estimated that out of 501.76 million tonnes of crop residues generated, 140.84 million tonnes crop residue was surplus in 2009 and 91.25 million tonnes crop residues were burnt.

Currently, 80% of the rice residue in the form of stubble is burnt in the field in the month of October and first two weeks of November. The tradition of burning of crop residues in Punjab started from the decade of the 1980s with the mechanization of the harvesting process. It was accepted by farmers because of low cost and no investment. There are also some additional reasons of the crop residue burning in the field such as shortage of laborers, shortage of time between harvesting of rice and wheat sowing, unavailability of residue buyers, etc. (Kaur 2017). However, in doing so, farmers ignored the impact of this practice on health and air quality. During these 3–4 weeks in the month of October–November, Delhi NCR (National Capital Region) chokes on smoke due to the burning of the paddy straw from the neighboring states (CPCB 2017).

Rice straw burning produces many harmful gases such as carbon dioxide, carbon monoxide, methane, and nitrogen oxides (Gupta et al. 2004). It is a major environmental issue due to air pollution, green house gases (GHG) emissions, health hazards, loss of nutrients of the soil, etc. (Pathak et al. 2006). The different sources of pollution in Delhi are vehicle/transport (39.1%), industry (25.4%), residential (5.7%), dust (18.1%), stubble and other biomass burning (11.7%). The contribution of stubble burning in the pollution in Delhi reached up to 44% on October 31, 2019 (Hindustan 2019 based on the research of SAFAR 2018). Figure 1 shows the percentage contribution of the stubble burning in producing PM2.5 in Delhi NCR (SAFAR 2019).

There are mainly two types of pollutants from the crop residue burning: (1) gaseous air pollutants and (2) particulate air pollutants. Gaseous air pollutants include SO_2, NO_X, O_3,

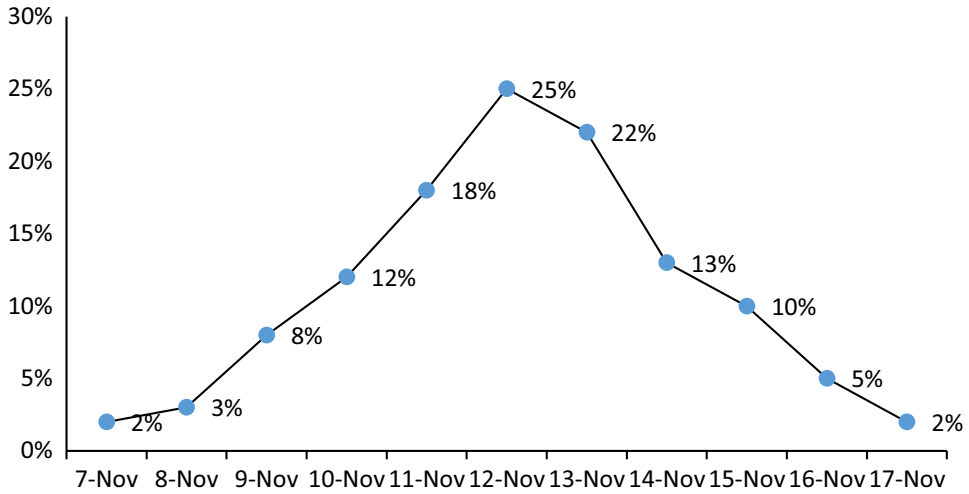


Fig. 1 Share of stubble burning in producing PM2.5 from November 7 to 17, 2019. *Source*: SAFAR (2019)

CO_2, CO, etc., whereas the particulate air pollutants include PM10, PM2.5, and PM1.0. If the size of the suspended particles in the air is less than 10 μm, it is known as PM10. If it is less than 2.5 μm, it is known as PM2.5. Similarly, if the suspended particle size is very fine (less than 1 μm), it is known as PM1.0 (CPCB 2016). These products of residue combustion are hazardous to the human health. Thus, some alternate solutions are required to prevent the open burning of the crop residue to save the environment and life of human being and animals.

Although farmers are aware of the environmental impact of the open burning of the crop residue in the field, they have no feasible and cost-effective options due to lack of infrastructure and technology. Lohan et al. (2018) followed a qualitative approach to analyze some possible options for disposition of crop residue. The present study has tried to adopt a data-based quantitative approach to analyzing possible substitutes for disposition of crop residue. A focus group discussion (FGD) has been used to get the opinion for the weights of different criteria to evaluate the substitutes for crop residues burning in the field. The substitutes for crop residue burning were rated on 5-point rating scale with the help of FGD members. Finally, the preference rankings of the substitutes based on the elimination and choice expressing reality (ELECTRE) analysis are produced considering all the constraints of the farmers. ELECTRE is an outranking method which outranks the unimportant substitutes, and ranks are provided to rest of the substitutes. If it is feasible to use all the substitutes with equal importance, it is good for the environment, but due to lack of technology and resources, all the substitutes of crop residue burning are not feasible. In this situation, preference ranking becomes more important for the farmers. Other multi-attribute utility techniques (MAUT) provide the ranks to all the substitutes without outranking the unimportant substitutes. Singh et al. (2017) used the fuzzy comprehensive method as a MAUT for assessment of the air quality in Harora river basin and took the readings of various sampling stations in west Tripura.

The rest of this paper has been arranged as follows: Sect. 2 presents the literature review of the issues of crop residue burning in India, Sect. 3 presents the research methodology, Sect. 4 presents analysis and findings, and Sect. 5 concludes the research findings.

2 Literature review

Crop residue burning is not a problem of India only, but it is also equally important for other countries like China, Russia, Thailand, Italy, and Pakistan. In China, many pollutants are released because of crop residue burning in the field, resulting in serious pollution of ambient air. Suqian with 4523 km^2 of total area under cultivation was selected as a case to be studied (Yang et al. 2008). Oanh et al. (2018) have done the research on emission of toxic air pollutants and greenhouse gases from crop residue open burning in Southeast Asia. Many research works have been published in the past on measurement of emissions due to the burning of crop residues and other biomass burning. Rice and wheat have main contributions in cereal crop residues in India. It is approximated as 44 million tons of rice and 24.5 million tons of wheat out of 84 million tons of cereal crop residues (IARI 2012). According to Ravindra et al. (2018), crop residue burning emissions will increase by 45% in 2050 considering 2017 as the base year. It has been observed that a thick blanket of smoke is formed over the northwestern states of India and extends up to the National Capital Region Delhi just at the end of the month of October. October–November is the season for the paddy stubble burning in the NCR Delhi (CPCB 2017).

Post-monsoon burning of the rice straw continues for several weeks, which increases the aerosol formation over the Indo-Gangetic Plains. However, pre-monsoon burning of the wheat stubble is not much harm to the atmosphere (Vadrevu et al. 2011; Singh and Kaskaoutis, 2014). The cloud of fog formation in the region of Indo-Gangetic Basin can be observed just after the end of the monsoon. The thickness of the aerosol layers may vary from the earth's surface to 4–4.5 km altitude having highest concentration very close to the earth's surface (Mishra and Shibata 2012). The smoke due to the stubble burning is very dangerous for the human being. China has also observed the severe effect of air pollution on the premature deaths (Liu et al. 2018). Approximately, 6,00,000 premature deaths per year due to poor air condition has been recorded in India (Lelieveld et al. 2015; Ghude et al. 2016; WHO 2016). The loss of average life expectancy in Delhi is approximately 6.4 years; however, it is recorded as 3.4 years for the rest of the country due to exposure of PM2.5 (Ghude et al. 2016).

2.1 Degradation of air quality due to crop residues burning

Air quality has significant variation throughout the year in the region of Delhi NCR (refer Fig. 4 given in Appendix 1). It has been observed that during the month of May (pre-monsoon) and October-November (post-monsoon) the air quality index becomes very poor. Delhi's pre-monsoon airshed (the term "airshed" is used to refer the spatial zone around the city where emissions can contribute to pollution within the city due to outdoor fires) covers mainly the croplands of Haryana and southern Punjab and post-monsoon airshed covers Punjab, Haryana, and across the border of Pakistan. The post-monsoon airshed averages mainly cropland (96%), whereas the pre-monsoon airshed averages cropland (83%) with some parts of the grassland (14%) (Tianjia et al. 2018). Figure 2 shows the air quality trends of Delhi from 2009 to 2015.

Burning crop residue in the field is an uncontrolled combustion due to lack of oxygen. During a small time period (in weeks), a very large area of the farms in Punjab is set into the fire and the availability of oxygen for complete combustion of the crop residue becomes insufficient. The major products of the incomplete combustion of stubble are carbon monoxide and unburnt carbon, but it also produces carbon dioxide, methane, nitrogen oxides, and small fraction of sulfur dioxide (Galanter et al. 2000; Hays et al. 2005). The PM2.5 stays in the air for a long time due to light weight and travels a long distance (hundreds of miles) in the direction of the wind (Singh and Panigrahy 2011; Jain 2016). The variations in the level of PM2.5 and PM10 from October 27, 2017, to November 30, 2017, in Delhi NCR due to the burning of paddy straw are very high. It has been observed that the

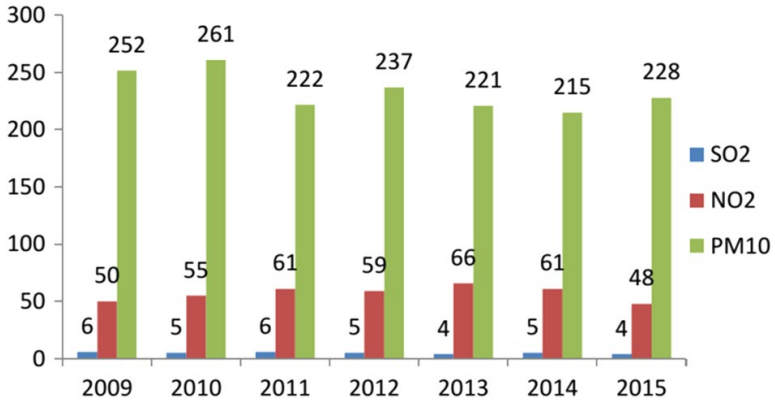


Fig. 2 Air quality trends in Delhi. *Source*: CPCB (2015)

Table 1 Area under major cereal crops and sugarcane, 2012–2013. *Source*: DAC (2014)

S. no	State	Area under major cereal crops and sugarcane (in thousands Hectares)			
		Rice	Wheat	Sugarcane	Crops residue prone to burning
1	Haryana	1215	2497	101	Rice, wheat, sugarcane
2	Punjab	2845	3512	83	Rice, wheat, sugarcane
3	Uttar Pradesh	5861	9734	2212	Rice, wheat, sugarcane
4	Rajasthan	125.6	3063.2	5.5	Minor cases of burning

Table 2 Production of crop residue (in Tonnes), 2014–2015. *Source*: Devi et al. (2017)

S. no	State	Rice	Wheat	Coarse cereal	Pulse	Oilseed	Sugarcane	Cotton	Jute and Mesta	Total
1	Haryana	4.7	14.7	1.4	0.1	0.8	3.0	0.8	0	25.4
2	Punjab	13	19.6	0.9	0.1	0.1	2.8	0.5	0	36.9
3	Uttar Pradesh	0.4	12.2	12.9	3	5.8	0.2	0.5	0	35.1
4	Rajasthan	14.3	31.3	6.1	2.2	0.9	54.4	0	0	109.2

maximum levels of PM2.5 and PM10 are ranging from November 7 to 14, 2017 (CPCB 2017).

This is the reason that the burning of crop residues in Punjab, Haryana, Rajasthan, and the farms in the border area of Pakistan chokes the regions of Delhi NCR chokes during pre- and post-monsoon. Table 1 shows the area of the land used for the rice, wheat, and sugarcane production for the states Haryana, Punjab, Uttar Pradesh, and Rajasthan. Table 2 shows the crop residue production from different crops in Haryana, Punjab, Uttar Pradesh, and Rajasthan. Table 3 shows the total crop residue production, surplus amount, and the amount of the crop residue burnt in the field.

Punjab is a big producer of paddy straw (23Mt) and wheat straw (17Mt) (IARI 2012; Jain et al. 2014; NAAS 2012; Chauhan 2012; Singh et al. 2008). Only 20% of the paddy straw is utilized as fodder for cattle, and other purposes and rests are burnt in the field. However, more than 80% of the wheat straw is used as cattle fodder and rest 20% is burnt in the field. The farmers in Haryana also burn the crop residues (approx. 9.08 Mt) in the field (Pandey 2019). The contributions of straw and husk of paddy and wheat among total crop residues are 85.6% and 78.48% in Punjab and Haryana, respectively. Total production of the crop residues in Punjab and Haryana is 50.75 Mt and 22.83 Mt, respectively,

Table 3 Amount of surplus and burned crop residues 2009. *Source*: DAC (2014)

S. no	State	Residue generated (Mt)	Residue surplus (Mt)	Residue burned (Mt)
1	Haryana	27.83	11.92	9.08
2	Punjab	50.75	24.83	19.65
3	Uttar Pradesh	59.97	13.53	21.92
4	Rajasthan	29.32	8.52	1.78

in 2009. The surplus amount of the crop residue in Punjab and Haryana is 24.83 Mt and 11.22 Mt, respectively. The crop residue burnt in Punjab and Haryana is 19.65 Mt and 9.08 Mt, respectively. 22.32 Mt is a surplus (Jain et al. 2014; NAAS 2012; Chauhan 2012; MNRE 2009; Sangeet 2016; Chauhan 2011).

2.2 The factors forcing the farmers to burn the crop residues

In Punjab, the practice of mechanized harvesting has been increased in the past two to three decades which avoids the use of costly manual labors. With combined harvesters, stubble is left in the field during the harvesting. It is costly to remove the stubble from the farm. The farmer finds it easy to burn the stubble for preparation of the field for the next crop (Gadde et al. 2009; Kumar et al. 2015). Some of the observed factors responsible for the burning of crop residues are discussed below.

2.2.1 Changing trend of manual to mechanical harvesting

Punjab and Haryana states have been shifted to the mechanized forming since last 2–3 decades due to increasing labor cost and a trend of labor to shift from agriculture to industry due to high wage in industry compared to agriculture. In Punjab, the involvement of manual laborers has been decreased tremendously from 62.67 to 35.96% during the last 40 years (from 1970–1971 to 2012–2013) (Lohan et al. 2015a). The popularity of mechanized farming has increased due to unavailability of manual labor and high wages (Lohan et al. 2018; Singh et al. 2011; Sidhu et al. 2015; Jat et al. 2009, 2011).

2.2.2 Use of combine harvesters

In Punjab, the cropping intensity has sharply increased due to increase in farm power in the last 40–50 years (Kumar et al. 2015). The farm power has reached to 5.68 kW ha^{-1} during these 4–5 decades. The combine harvester has eliminated the requirement of a large manpower in cultivation. The combine harvester leaves a huge mass of scattered residues in the field in the form of stubble (Mittal et al. 2009). Thus, the burning of the stubble is an easy and economical method for the farmers to prepare the farm for the timely sowing of wheat.

2.2.3 Lack of time to prepare the field for the next crop

The time gap between the harvesting of rice and the sowing of the wheat is very small. It has been observed that the time gap is 7–10 days between the harvesting of the Basmati rice and sowing of the wheat and 15–20 days between the harvesting of coarse grain rice and sowing of the wheat (Gupta 2011). Approximately, 7.5 tonnes/ha of the straw remains in the field after harvesting of paddy; farmers used to burn these residues in the field (Singh and Panigrahy 2011; Jain et al. 2014).

2.2.4 Concept of burning of weeds/pests with crop residues

The farmers believe that the burning of crop residues in the field leads to burning weeds, insects, diseases, and pests (Jain et al. 2014; Gupta 2011). Also, the residual ash works as fertilizer due to the presence of potassium in the ash. It may raise the short-term gain due

to availability of specific nutrients like phosphorus and potassium, but it may cause a loss of other nutrients and organic carbon (Jat et al. 2009; Singh et al. 2006).

2.3 Observable impacts of crop residue burning

The life expectancy of the people in Delhi NCR is being adversely affected. People are suffering from chronic diseases and respiratory problems. The crop residue burning has an additional contribution with some other factors responsible for pollution, such as industry, vehicle, dust, and residential burning of fuel. Some of the observed consequences of the crop residue burning are discussed below.

2.3.1 Degradation of air quality and increasing the concentration of greenhouse gases

Crop residue burning degrades the air quality due to emission of various gases, including gases responsible for global warming (Ramanathan and Carmichael 2008; Ioannidou and Zabaniotou 2007).

2.3.2 Releasing the soot particles in the atmosphere causes the smog formation

Crop residue burning leads to production of suspended particulate matter in the atmosphere in the form of PM2.5 and PM10. PM2.5 floats in the air for a long time due to lightweight and creates the cloudy type of the atmosphere due to smog formation. It spreads in a large area in the direction of wind. Van den Elshout et al. (2014) used common air quality index to monitor PM2.5 for 31 urban backgrounds and 27 street stations all across Europe in 2006.

2.3.3 Hazardous to the health of human and other animals

The emission from burning of crop residues also causes severe impacts on human health. It may cause the people suffering from chronic disease related to respiratory system, heart and lung problems, blood, eye, and skin irritation (Singh et al. 2009, 2011; Mittal et al. 2009; Pathak et al. 2010; Long et al. 1998; Treets et al. 2003).

2.3.4 Loss of the soil fertility

The stubble burning causes the loss of the nutrients present in the farm. It also leads to evaporation of the elements such as carbon, nitrogen, phosphorus, potassium, and sulfur, which are already present in the various crops (Lefroy et al. 1994). It has been approximated that the losses of nitrogen, phosphorus, and potassium are 79.38 kg, 183.71 kg, and 108.86 kg per hectare, respectively, due to stubble burning (Jat et al. 2013; Teodoro and Mendoza 2016; Yang et al. 2003; Singh and Sidhu 2014). Apart from changes in chemical properties and loss of nutrients, it also affects the biochemical properties, i.e., soil enzymes (Amador et al. 1997). Soil enzymes play a vital role in the catalysis of those reactions that are necessary for organic matter decomposition and cycling of nutrients. It is also involved in energy transfer, crop productivity, and environmental quality. The soil enzymes are important indicators of soil health in terms of soil-friendly microbes, worms, insects, etc. (Tabatabai 1994). These worms and microbes are also burnt with the crop residue.

2.4 Alternative uses of crop residues

The paddy and wheat straw are very useful. A large part of the wheat straw is used as the cattle fodder, and only a small part of the wheat straw in the form of stubble is burnt in the field. Based on literature review, twelve substitutes for the crop residues burning are identified as discussed in following section.

2.4.1 Amalgamation of stubble in soil/in situ incorporation/ back in soil

Amalgamation of stubbles in soil is a good substitute for the burning of paddy residue (Sidhu and Beri 2005). The incorporation of paddy residue in the soil increases the organic carbon content by 14–29% (Singh et al. 1996), but the minimum time between the amalgamation and sowing of the wheat must be at least 3 weeks. But, if the same stubble is amalgamated in the soil only a few days before the sowing of the wheat, it arrests the inert nitrogen available in the soil and results in a nitrogen deficit (Singh et al. 1992). In situ incorporation decomposes the residues to some nutrients and also enhances the pH value, organic carbon, infiltration rate, and water holding capacity of the soil (Gupta et al. 2004).

2.4.2 Crop stubble as animal fodder/cattle shed and bed preparation

One of the substitutes for the paddy straw burning may be fodder, bed, and preparation of shed of the cattle. The paddy straw is normally not used as fodder for cattle due to the presence of high silica and lingo-cellulosic content which is not easily digestible except straw of Basmati rice and wheat straw (Jain 2016; Kumar et al. 2015). Wheat straw is frequently used as cattle fodder. Basmati rice straw is mostly used as cattle fodder throughout the country. However, 7–10% of the Basmati Paddy area are also reported to be affected by the incidence of crop residue burning (DAC 2019). The paddy straw is frequently used as bed and shed for the cattle in the winter, which protects the cattle from extreme cold and prevents the chances of injury and lameness (Kumar et al. 2015).

2.4.3 Wheat and paddy straw in mushroom farming

Paddy straw is a key ingredient for the cultivation of mushroom in Punjab (Choudhary et al. 2009), but farmers generally use wheat straw in place of paddy straw. For the production of mushroom, wheat straw is first washed and dried. After drying of the wheat straw, it is cut properly and bundles of straw are prepared.

2.4.4 Rice straw mulching

Mulching of Rice straw is a process to conserve the moisture of the soil. It may be used for the crops like wheat, maize, sugarcane, sunflower, soybean, potato, and chilli production since these crops require wetland (Singh et al. 2010; Sekhon et al. 2008; Arora et al. 2011). It also maintains the minimum soil temperature required in the winter and gives shading effect in the summer. Paddy straw is also used for mulching purpose in Haryana enormously. The Governments of Haryana and Punjab have proposed under the annual work plan of RKVY (Rashtriya Krishi Vikas Yojana) for seeking central assistance for providing machineries (happy seeder/turbo seeder/shredder/ baling machines and

zero-seed-cum-fertilizer drill) to the farmers to facilitate in situ management of crop residue and retaining the straw as surface mulching (DAC 2014).

2.4.5 Back in soil after composting

Composting is the decomposition of rice straw to enable recovery of portions of its nutrients and organic components if the feedstock materials have a high nitrogen content to obtain a better carbon-to-nitrogen ratio. Rice straw takes a longer time to decompose it may take a year. Scientists have already developed a rapid composting technique to convert huge piles of rice straw into the organic-rich soil. Generally, it takes about 45 days to prepare this rice straw compost.

2.4.6 Domestic fuel

The people of the rural areas in some states of Jammu & Kashmir, Uttarakhand, and Himachal Pradesh are dependent on fuel wood (combustion with dung cake/ wood/ coal) for cooking (Jain et al. 2014; Kumar et al. 2015). But now things have changed and the main reason for the unpreferred use of rice straw is the difficulty in procurement due to lightweight and occupying more volume.

2.4.7 Fuel for power plants

Few power plants are established in Punjab, which are using rice straw as fuel for power production. Fly ash and bottom ash as the by-products of fuel combustion in power plant may be used in the cement and brick manufacturing industry. The paddy straw may be used in the form of bales directly in the furnace or in the form of shredded straw with pulverized coal. The handling and storage of straw in the form bales become easy.

2.4.8 In biogas and biochar generation/bioethanol

The anaerobic digestion, gasification, and pyrolysis technologies are used to produce biogas and other forms of energy from the biomass (Koopman and Koppejan 1997; IARI 2012; Lohan et al. 2015b). It also reduces the concentration of CO_2 in the environment (Brar et al. 2000; Lohan et al. 2012; Dhaliwal et al. 2011). One tonne dry crop residue can produce 250–350 L of ethanol. It may help in replacing the use of gasoline up to some extent (Jeffery et al. 2011).

2.4.9 Making pellets

The crop residue may be used as a fuel in the pellet form. The pellet mill is used to crush, press, compact and form the straw, peanut shell, cob, cotton bar, soybean rod, weeds branches, leaves, sawdust, bark, and other solid wastes to prepare the pellets. This kind of fuel has high efficiency and is easy to store. It can also be used as main fuel for the industrial boiler (Verma 2014).

2.4.10 Raw materials for paper and pulp industry

A pulping technology could eliminate waste by turning rice straw into the paper. The best method extracts cellulose from the straw to make paper and natural phenolic materials. Thus, the major portion of the burning paddy straw can be utilized as pulp for paper and cardboard.

2.4.11 Mixing with plastics

The paddy straw can also be used as reinforcing the material in plastics. The paddy straw is shredded into small pieces of 1.5–2 mm size and mixed with polypropylene. The mixture is then extruded into granules using a twin screw extruder (Verma 2014). The various substitutes for the crop residues burning are summarized in Table 4.

2.4.12 Composite materials

In India, rice husk, jute, stalk, and coconut fibre are produced in large amount. All these natural fibres have excellent physical and mechanical properties and can be utilized more effectively in the development of composite materials for various applications like hard board (Pandey and Sujatha 2011).

All the alternative solutions to the crop residue burning, their pros and cons with references are summarized in Table 4.

2.5 Criteria to evaluate the solutions for crop residue burning

The various substitutes for the crop residue burning have already been proposed by the researchers and industry experts. The main issue is to find the feasibility of application of these substitutes as per the opinion of the farmers. Some of the important criteria to rate these substitutes are the economy of the residue management, the time required for composting the residues, air pollution, soil fertility, and several feasible domestic and industrial uses of the crop residues. All the substitutes may not be feasible to be applied in all the environmental/economic conditions. There may be many constraints with the application of some the substitutes such as logistic costs of the paddy straw from the farm to the power plant or paper and pulp company, economy, time, etc. Therefore, in this study, some of the leading factors/criteria have been identified (as summarized in Table 5) from the literature review and with consultation with farmers/experts based on which the substitutes for the crop residues burning are rated.

Reviewing the past literature, it is observed that most of the studies are based on the reason of crop residue burning, its impact, and alternative solutions. Very few studies are based on the opinion of the farmers for the alternative usage of the crop residue. Due to lack of technology, resources, logistics and storage and time constraints, many alternative solutions may not be feasible for all the farmers. Thus, based on the opinion of the farmers a priority rank of these solutions is needed. In this study, we have produced the ranks for the different alternative solution to the crop residue burning.

Table 4 Various substitutes for the crop residues burning in the farm

S. no	Substitutes for the crop residues burning	Pros	Cons	References
1	Amalgamation of stubble in soil/in situ incorporation (S1)	The incorporation of paddy residue in the soil increases the organic carbon content by 14–29%	Minimum time required or composting with application of microbial is 3 weeks and without microbial application 90 days	Lohan et al. (2018), Sidhu and Beri (2005), Singh et al. (2005), Kumar et al. (2015)
2	Animal fodder/shed/bed preparation (S2)	Only wheat straw (Bhusa) is used as fodder for the cattle	The paddy straw is normally not used as fodder for cattle due to the presence of high silica and lingocellulosic content which is not easily digestible except straw of Basmati rice and wheat straw	Lohan et al. (2018), IARI (2012), Kumar et al. (2015), Kumar et al. (2013)
3	Mushroom farming (S3)	Suitable for the mushroom farming and solving the scarcity of vegetable	The short life cycle does not allow to produce mushroom farming at large scale	Lohan et al. (2018), Choudhary et al. (2009), Roy and Kaur (2016)
4	Rice straw mulching (S4)	It conserves the moisture of the soil	It may absorb the available nitrogen in the soil and may result in decreasing the yield of the crops	Lohan et al. (2018), Jain et al. (2014), Beri et al. (1992)
5	Back in the soil after composting (S5)	It is one of the important methods to improve the soil fertility	Time of composing is long, and the straw is to be collected in one place for composting	Douglas and Rickman (1992), Kumar and Goh (1999)
6	Domestic fuel (S6)	It solves the shortage of fuel/energy for cooking	It is not good for the health and environment	Lohan et al. (2018), Jain et al. (2014), Kumar et al. (2015)
7	Fuel for the power plant (S7)	It reduces the use of coal in thermal power plant	Storing and transportation is difficult due to bulky in nature	Kumar et al. (2015), Cheewaphongphan et al. (2018)
8	Biogas, biochar, and bioethanol preparation (S8)	It is good for producing the biofuel and replacing the fossil fuels	Lack of resources and facilities to produce the bioenergy using stubble	IARI (2012), Brar et al. (2000), Lohan et al. (2012)
9	Making Pellets (S9)	It reduces the space required to store the stubble in the form of pallets and also increases the calorific value due to compaction	There is lake of sufficient facilities for making pallets from the stubble and rice husk	Verma (2014), Karkania et al.(2012), Sultana et al. (2010)

Table 4 (continued)

S. no	Substitutes for the crop residues burning	Pros	Cons	References
10	Mixing with plastics (S10)	It is good for the industrial application in plastic and ceramic industry	In India, it is not popular method. It is in the experimental phase in China	Verma (2014)
11	Preparation of hardboard for furniture industry (S11)	It is an important application in hardboard manufacturing	The strength and life of the hardboard is less than the wooden hardboard	Lohan et al. (2018)
12	Paper and pulp industry (S12)	It may be used in the pulp industry and low-quality paper manufacturing	Lack of resources to extract the cellulose from the rice stubble to make the paper	Lohan et al. (2018), Kaur et al. (2017)

Table 5 Criteria used for ranking of substitutes for the crop residue burning

Criteria	References
1. Cost of residue management (C1)	Sultana et al. (2010), Brechbill and Tyner (2008), Campbell (2007), Kumar et al. (2003), Liu (2008)
2. Time of residue composting/processing (C2)	Lohan et al. (2018), Singh et al. (1992), Gupta (2011)
3. Air quality degradation (C3)	Lohan et al. (2018), Oanh et al. (2011), Zhang et al. (2008), Sahai et al. (2007)
4. Loss of soil fertility (C4)	Singh et al. (2010), Gupta et al. (2004), Lohan et al. (2018), Singh et al. (2004)
5. In-house application (C5)	Lohan et al. (2018), Jain et al. (2014), Kumar et al. (2015)
6. Industrial applications (C6)	Lohan et al. (2018), Verma (2014)
7. Need of logistics facilities (C7)	Lohan et al. (2018), Sultana et al. (2010)
8. Bioenergy and power production (C8)	IARI (2012), Koopman (1997), Lohan et al. (2015b)
9. Govt. regulations to straw burning (C9)	Lohan et al. (2018), PSCST-TERI (2014), NDTV (2016)

3 Research methodology

In this study, focus group discussion (FGD) is used to gather the information for the quantitative analysis. Analytic hierarchy process (AHP) is considered for finding the priority weights of criteria to evaluate the substitutes for the crop residue burning. The pairwise comparison of the criteria is based on the most common opinion of the farmers. After finding the priority weight of the criteria, various substitutes for the crop residue burning are rated with respect to each criterion on a five-point rating scale and a decision matrix is prepared. Elimination Et Choix Traduisant la REalité (ELECTRE) is used as a method of outranking to find the rank of the different substitutes for the crop residue burning. The detailed procedures of AHP and ELECTRE have been mentioned in the following subsections.

3.1 Focus group discussion (FGD)

FGD is a way to get the collective opinion of the people concerned with the topic of interest. It is a type of group decision making. In this process, the group is guided by a moderator or facilitator and introduces the topics for discussion. The main strength of FGD relies on agree or disagree of participants with each other so that it provides an insight into how a group thinks about the issue, about the range of opinion and ideas. FGDs can be used to explore the meanings of survey findings that cannot be explained statistically, the range of opinions/views on a topic of interest and to collect a wide variety of local terms (Bloor et al. 2001; Stewart and Shamdasani 1990). The main point of discussion and the opinion of the FGD participants are summarized in Table 6.

In this study, five farmers, two academic experts, and two industry experts working in the agro-based industry (total nine people) are requested to participate in the focus group discussion. The various criteria for evaluation of the solutions for crop residue burning were introduced with different solutions for crop residue burning. They were asked to produce the merit and demerit of different criteria and the solutions for crop residue burning. They were also asked to do the pairwise comparison of the shortlisted criteria to find the

Table 6 Point of discussion and opinion of the participants

S. no	Points of discussion	Opinion of participants
1	Contribution of crop residue burning in degradation of air quality of Delhi NCR during the peak season, especially in the month of October–November	Farmers accepted that contribution of the stubble burning increases the PM 2.5 level in the air of Delhi NCR. But, they also suggested that the contribution of vehicle and the industry are also very high in producing the pollution in Delhi NCR
2	The geographic situation of Delhi NCR also plays a role in the formation of gas chamber in the large region as it is mentioned in many news article	Academic and industrial participants gave the emphasis on the geographical reason. They said that during the peak season wind flows from north to south which carries the smoke of the residue burning from Punjab and Haryana towards Delhi NCR, but the wind speed becomes very slow in the Delhi NCR and forms a blanket of smoke. They also highlighted that the level of pollution becomes low suddenly even in the peak season after raining and increasing the wind speed. Thus, wind speed and direction also plays an important role
3	Role super harvester in minimizing the residue burning	Super harvester is a costly equipment, and it is out of reach from the general people. About two-thirds of the farmers have less than 5 acre land for cultivation
4	Provision of 50% and 80% subsidies for the agro machine to the individual farmers and cooperative societies or the group of farmers, respectively	Framers opined that this is under planning and has not reached to the farmers on implementation level
5	Main reason of stubble burning of rice	Farmers told only 20–25 days are available between the harvesting of rice and sowing of the wheat which compels us to burn the stubble of rice
6	Mixing the stubble in the soil using different methods	Farmers opined that the time required for composting without any microbial treatment is 3 months, and with treatment of microbial, it takes more than 1 month. If it is mixed just few days before into the soil, it decreases the yield of wheat
7	Use of the residue as a cattle fodder	According to the farmers, we consume most of the wheat residue as the cattle feeder, but we are not using the rice straw as a cattle feeder due to less digestive and less nutritive for the cattle
8	Supply of stubble as fuel to the power plant	Power plant is not sufficient to consume all the rice residue and also due to bulky in nature logistics cost is high
9	Use of residue as bioenergy	Resources are not available to consume the rice stubble to produce the bioenergy as a biogas or bioethanol
10	Others use as paper industry, plastic industry, etc.	Farmers are facing the lack of resources and support to supply the same to the paper and pulp industry or plastic industry. Plastic industry also facing the lack of technology to consume the rice residue to produce the microsilica which is used in the plastic materials

Table 6 (continued)

S. no	Points of discussion	Opinion of participants
11	Evaluation of the alternative solution for crop residue burning using multiple criteria	Farmers are asked to rate the various criteria produced them like time of processing, cost, air quality degradation, etc. Pairwise comparison was done with the help of coordinator of the FGD

weights. They were asked to rate the solutions of crop residues burning on 5-point rating scale from least feasible to most feasible considering all the criteria one by one. The final decision taken by the FGD members is incorporated as inputs for AHP and ELECTRE analysis.

The various substitutes for the crop residues burning are collected from the literature review, the current practices, and the researches going on. The criteria for evaluating the substitutes are finalized after consultation with the members of FGD. Some of the practical problems like long composting time, the lack of infrastructure, high logistics cost, lack of storage or dumping ground, lack of industrial setup, etc., have been observed. Tables 1, 2, 3, 4 show the list of the substitutes for the burning of crop residues burning and the criteria to compare the substitutes.

3.2 Analytic hierarchy process (AHP)

AHP is a tool to find the rank of decision alternatives. Here, criteria for rating the substitutes for the crop residues burning are ranked using AHP. These criteria are compared with each other on the 9-point rating scale. Saaty (1980) developed the concept of AHP as a decision-making tool. In this process, the various criteria and sub-criteria with some alternatives may be shown in different levels of a hierarchy as per the requirement. The steps used to find the importance rating of criteria using AHP is discussed below:

Step 1 To prepare a pairwise comparison matrix, each criterion is compared with respect to each other on the 9-point rating scale. If there are 'n' criteria, a pairwise comparison matrix of $n \times n$ is developed, where 'n' is the number of criteria to evaluate the alternatives of crop residues burning.
Step 2 After preparation of the pairwise comparison matrix, each element (a_{ij}) of a column is divided by the summation of all the elements of the same column to standardize the element.
Step 3 Average value of each row of matrix $A^{'}$ (i.e., the matrix produced in step 2) is calculated.
Step 4 The consistency of the decision taken for pairwise comparison in decision matrix is checked.

After finding the importance of various criteria, the ratings of the substitutes for the crop residue burning are rated using ELECTRE.

3.3 Elimination and choice expressing reality (ELECTRE)

Bernard Roy proposed this method at SEMA Consultancy Company. It was the first time applied in 1965, but it was soon applied for choosing, ranking, and sorting of alternatives. A wider application of ELECTRE has been reported in the past research work in its various forms, including its fuzzification (Kumar et al. 2016, 2017a, b). In this paper, it is used to rank the substitutes for the crop resides burning. ELECTRE outranks the unimportant alternatives and ranks the rest of the alternatives considered in the analysis. But other MAUTs (Multi-Atrribute Unitity Techniques) rank all the alternatives without outranking the unimportant alternatives. This is the uniqueness of the method "ELECTRE."

The detailed procedures for the application of the ELECTRE (Shanian and Savadogo 2006; Hashemi et al. 2014) are mentioned below:

All the substitutes for the crop residues burning are rated with respect to the various criteria on a 5-point rating scale and shown as a decision matrix (Eq. 1):

$$X = \left[x_{ij}\right] \tag{1}$$

Normalize the decision matrix. The normalization can be done using Eq. (2):

$$R = \left[r_{ij}\right] \tag{2}$$

where $r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{m} x_{ij}^2}}$

$i = 1, 2, \ldots n$ is the number of criteria; and $j = 1, 2, 3, \ldots, m$ is the number of substitutes.

Find the weighted normalized decision matrix by multiplying the normalized element of the matrix with corresponding weights of the criteria (as found using AHP or some other methods) using Eq. (3).

$$V = [v_{ij}] \tag{3}$$

where $v_{ij} = r_{ij} \times w_i$.

Determine the concordance and discordance sets of criteria using the conditions shown in Eqs. (4) and (5)

$$C_{kl} = \left\{I | v_{ik} \geq v_{il}\right\} \tag{4}$$

$$D_{kl} = \left\{I | v_{ik} < v_{il}\right\} = I - C_{kl} \tag{5}$$

where C_{kl} is the set of concordance criteria for which the values of v_{ij} of kth substitute is greater or equal to the values of v_{ij} of lth substitute. Similarly, D_{kl} is the set of discordance criteria for which the values of v_{ij} of kth substitute are less than the values of v_{ij} of lth substitute.

Prepare the concordance matrix using Eq. (6):

$$\begin{aligned} &C = \left[c_{kl}\right], \text{and} \\ &c_{kl} = \sum_{i \in C_{kl}} w_i / \sum_{i=1}^{n} w_i; \quad \text{where} \quad 0 \leq c_{kl} \leq 1 \end{aligned} \tag{6}$$

Here, c_{kl} is a summation of the weights of concordance criteria for which kth substitute is preferred over the lth substitute.

Prepare the discordance matrix using Eq. (7):

$$\begin{aligned} &D = \left[d_{kl}\right], \text{and} \\ &d_{kl} = \frac{\max\limits_{i \in D_{kl}} |v_{ik} - v_{il}|}{\max\limits_{i \in I} |v_{ik} - v_{il}|}; \quad \text{where} \quad 0 \leq d_{kl} \leq 1 \end{aligned} \tag{7}$$

Here, the numerator shows the maximum difference (numerically) between the weighted normalized elements of kth and lth substitutes for the criteria for which kth substitute is

inferior to the lth substitute. Denominator shows the maximum difference (numerically) between kth and lth substitutes for all the criteria.
Find the average value of c_{kl} using Eq. (8)

$$\bar{c} = \sum_{k=1,k\neq l}^{m} \sum_{l=1,l\neq k}^{m} c_{kl}/m(m-1) \tag{8}$$

Prepare a Boolean matrix E and put the value as shown in Eq. (9):

$$\begin{array}{ll} e_{kl} = 1; & \text{for } c_{kl} \geq \bar{c} \\ 0; & \text{for } c_{kl} < \bar{c} \end{array} \tag{9}$$

Find the average value of d_{kl} using Eq. (10)

$$\bar{d} = \sum_{k=1,k\neq l}^{m} \sum_{l=1,l\neq k}^{m} d_{kl}/m(m-1) \tag{10}$$

Prepare a Boolean matrix E and put the value as shown in Eq. (11):

$$\begin{array}{ll} f_{kl} = 0; & \text{for } d_{kl} \geq \bar{d} \\ 1; & \text{for } d_{kl} < \bar{d} \end{array} \tag{11}$$

Find the aggregate dominance matrix G by multiplying the elements of Boolean matrix E with the corresponding elements of Boolean matrix F as Eq. (12):

$$\begin{array}{l} G = \left[g_{kl}\right] \\ \text{where} \quad g_{kl} = e_{kl} \times f_{kl} \end{array} \tag{12}$$

On the basis of the final dominance matrix, we can find the dominance of one alternative over another and we can map the outranking relationship of the substitutes with respect to the other dominant substitutes.

4 Result discussion

The complete analysis is based on the opinion of the members of the focussed group discussion (FGD). The common opinion obtained from the discussion of the FGD members. Total 9 members were involved in the focus group discussion, out of which five farmers, two academic experts from environmental engineering, and two experts from industry were called. We have not used a general questionnaire like a market survey as a method of inputs. We used focused group discussion and simple questionnaire to know the opinion and view of the farmers for residue burning. Generally, FGD involves group interviewing in which a small group of usually 8–12 people. It is led by a moderator (interviewer) in a loosely structured discussion of various topics of interest.

We had contacted the farmers through gram panchayat. We had identified the farmers who are involved in the cultivation of mainly rice and wheat in more than 10 acres of land. It is observed that approx. two-thirds of the farmers of Punjab have less than 5 acres land. Two farmers are contacted from Punjab, one from Haryana, and two from

western UP. After the first round of discussion, we facilitated them to come at Delhi for focused group discussion with our team of academicians and industry experts. On the basis of the discussion, their common opinion is incorporated in the study.

In place of directly using the questionnaire, the major concern with the crop residues burning and the objective of the study was explained by the authors as the moderator. Since most of the farmers' knowledge is not technically and scientifically sound, various pros and cons of the different substitutes were highlighted by the academic experts and the industry experts and solutions were rated with respect to the different criteria on the basis of feasibility. The 5-point rating scale (from 1 least feasible to 5 most feasible) was used.

After deciding the priority or weights of the criteria and rating of the different solutions for the crop residues burning in the farm, a decision matrix has been prepared (Eq. 1). The elements of the decision matrix are normalized to represent the elements of the matrix on the same scale (0–1). The normalization is done using Eq. (2). After normalization, the normalized elements are multiplied by the corresponding weights of the criteria (Eq. 3) to get the weighted normalized matrix. Table 7 shows the weighted normalized matrix.

From the weighted normalized matrix, concordance and discordance sets of criteria are decided using the conditions in Eqs. (4) and (5), respectively. For example, if we compare the columns S1 and S2 in the weighted normalized matrix, the values of v_{s1} are greater than or equal to the values of v_{S2} for the criteria C1, C4, C7, C6, C8, and C9. Therefore, the concordance set of criteria for the cell S1S2 is {1, 4, 6, 7, 8, 9}. Similarly, in the same columns, S1 and S2 in the weighted normalized matrix, the values of v_{s1} are less than the values of v_{s2} for the criteria C2, C3, and C5. Therefore, the discordance set of criteria for the cell S1S2 is {2, 3, 5}. From the concordance sets of criteria, the concordance matrix has been developed by summing up the weights of concordance criteria as shown in Eq. (6). For Example, the element of the cell S1S2 in concordance matrix shows the summation of the weights of the criteria C1, C4, C6, C7, C8, and C9 (0.163 + 0.062 + 0.067 + 0.127 + 0.048 + 0.069 = 0.536). Table 8 shows the concordance matrix.

The discordance matrix is developed using Eq. (7). For example, the element of the cell S1S2 in discordance matrix is calculated as:

$$= \frac{\max(|0.0352 - 0.0704|, |0.0282 - 0.0353|, |0.0037 - 0.0185|)}{\max(|v_{S1} - v_{S2}|)}$$

$$= \frac{0.0352}{0.0450} = 0.781$$

where

$$\max(v_{S1} - v_{S2}) = \max(|0.0750 - 0.0300|, |0.0352 - 0.0704|, |0.0282 - 0.0353|, |0.0323 - 0.0194|, |0.0037 - 0.0185|, |0.0054 - 0.0054|, |0.0748 - 0.0299|, |0.0058 - 0.0058|, |0.0208 - 0.0208|) = 0.0450$$

Here, in the numerator, only those cells of the Columns S1 and S2 are considered for which $v_{S1} - v_{S2} < 0$ and in denominator all the cells of S1 and S2 are considered to find the maximum numerical difference between v_{S1} and v_{S2}. Table 9 shows the discordance matrix.

From the concordance matrix, Boolean matrix E has been developed using Eqs. (8) and (9). $\bar{c}$ is calculated as $\bar{c} = \sum_{k=1,k\neq l}^{m} \sum_{l=1,l\neq k}^{m} c_{kl}/m(m-1) = \frac{72.339}{132} = 0.548$. If $\bar{c}$ is greater than or equal to c_{kl} the value in the Boolean matrix will be 1 otherwise 0. Table 10 shows the Boolean matrix E.

Similarly, from the discordance matrix, Boolean matrix F has been developed using Eqs. (10) and (11). $\bar{d}$ is calculated as $\bar{d} = \sum_{k=1,k\neq l}^{m} \sum_{l=1,l\neq k}^{m} d_{kl}/m(m-1) = \frac{89.637}{132} = 0.679$.

Table 7 Weighted normalized matrix

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
C1	0.0750	0.0300	0.0600	0.0600	0.0450	0.0300	0.0450	0.0450	0.0450	0.0300	0.0450	0.0300
C2	0.0352	0.0704	0.0352	0.0704	0.0352	0.0352	0.1408	0.1408	0.1408	0.0704	0.1408	0.1056
C3	0.0282	0.0353	0.0353	0.0282	0.0282	0.0071	0.0212	0.0212	0.0353	0.0353	0.0353	0.0212
C4	0.0323	0.0194	0.0065	0.0323	0.0323	0.0065	0.0065	0.0065	0.0065	0.0065	0.0065	0.0065
C5	0.0037	0.0185	0.0148	0.0037	0.0037	0.0185	0.0037	0.0037	0.0037	0.0037	0.0037	0.0037
C6	0.0054	0.0054	0.0054	0.0054	0.0054	0.0054	0.0268	0.0268	0.0268	0.0268	0.0268	0.0268
C7	0.0748	0.0299	0.0299	0.0599	0.0599	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150
C8	0.0058	0.0058	0.0058	0.0058	0.0058	0.0175	0.0291	0.0291	0.0058	0.0058	0.0058	0.0058
C9	0.0208	0.0208	0.0208	0.0208	0.0208	0.0125	0.0166	0.0208	0.0208	0.0208	0.0208	0.0208

Table 8 Concordance matrix

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	–	0.536	0.868	0.668	1	0.92	0.553	0.553	0.501	0.501	0.501	0.601
S2	0.648	–	0.837	0.648	0.648	0.952	0.39	0.541	0.438	0.933	0.438	0.601
S3	0.648	0.574	–	0.479	0.742	0.841	0.553	0.395	0.601	0.601	0.601	0.601
S4	0.71	0.868	0.868	–	1	0.89	0.553	0.304	0.501	0.833	0.501	0.601
S5	0.71	0.536	0.705	0.505	–	0.92	0.553	0.276	0.501	0.501	0.501	0.601
S6	0.479	0.31	0.509	0.147	0.479	–	0.221	0.184	0.269	0.432	0.269	0.432
S7	0.606	0.61	0.509	0.606	0.642	0.968	–	0.196	0.831	0.831	0.831	0.931
S8	0.548	0.679	0.578	0.548	0.711	0.968	1	–	0.9	0.9	0.9	1
S9	0.648	0.779	0.678	0.648	0.811	0.92	0.952	0.952	–	1	1	1
S10	0.648	0.779	0.678	0.648	0.648	0.92	0.457	0.457	0.505	–	0.505	0.668
S11	0.648	0.779	0.678	0.648	0.811	0.92	0.952	0.92	1	1	–	1
S12	0.548	0.679	0.578	0.548	0.548	0.92	0.457	0.457	0.405	0.9	0.405	–

Table 9 Discordance Matrix

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	–	0.781	0.246	1	0	0.246	1	1	1	0.587	1	1
S2	1	–	0.852	1	0.426	0.33	1	1	1	1	1	1
S3	1	1	–	1	1	0.387	1	1	1	1	1	1
S4	0.426	0.492	0.315	–	0	0.329	1	1	1	0.477	1	0.783
S5	1	1	0.501	1	–	0.329	1	1	1	0.783	1	1
S6	1	1	1	1	1	–	1	1	1	1	1	1
S7	0.567	0.212	0.142	0.637	0.425	0.14	–	1	0.605	0.2	0.605	0.118
S8	0.567	0.212	0.142	0.637	0.425	0.14	0	–	0.605	0.2	0.605	0
S9	0.567	0.212	0.142	0.637	0.425	0.14	1	1	–	0	0	0
S10	1	0.697	0.852	1	1	0.42	1	1	1	–	1	1
S11	0.567	0.212	0.142	0.637	0.425	0.14	1	1	0	0	–	0
S12	0.85	0.425	0.426	1	0.637	0.21	1	1	1	0.4	1	–

Table 10 Boolean Matrix E

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0	0	1	1	1	1	0	1	0	0	0	1
S2	1	0	1	1	1	1	0	0	0	1	0	1
S3	1	1	0	0	1	1	1	0	1	1	1	1
S4	1	1	1	0	1	1	1	0	0	1	0	1
S5	1	0	1	0	0	1	1	0	0	0	0	1
S6	0	0	0	0	0	0	0	0	0	0	0	1
S7	1	1	0	1	1	1	0	0	1	1	1	1
S8	1	1	1	1	1	1	1	0	1	1	1	1
S9	1	1	1	1	1	1	1	1	0	1	1	1
S10	1	1	1	1	1	1	0	0	1	0	0	1
S11	1	1	1	1	1	1	1	1	1	1	0	1
S12	1	1	1	1	1	1	0	0	0	1	0	0

Table 11 Boolean matrix F

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0	0	1	0	1	1	0	0	0	1	0	0
S2	0	0	0	0	1	1	0	0	0	0	0	0
S3	0	0	0	0	0	1	0	0	0	0	0	0
S4	1	1	1	0	1	1	0	0	0	1	0	0
S5	0	0	1	0	0	1	0	0	0	0	0	0
S6	0	0	0	0	0	0	0	0	0	0	0	0
S7	1	1	1	1	1	1	0	0	1	1	1	1
S8	1	1	1	1	1	1	1	0	1	1	1	1
S9	1	1	1	1	1	1	0	0	0	1	1	0
S10	0	0	0	0	0	1	0	0	0	0	0	0
S11	1	1	1	1	1	1	0	0	1	1	0	1
S12	0	1	1	0	1	1	0	0	0	1	0	0

Table 12 Dominance matrix G

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
S1	0	0	1	0	1	1	0	0	0	0	0	0
S2	0	0	0	0	1	1	0	0	0	0	0	0
S3	0	0	0	0	0	1	0	0	0	0	0	0
S4	1	1	1	0	1	1	0	0	0	1	1	1
S5	0	0	1	0	0	1	0	0	0	0	0	0
S6	0	0	0	0	0	0	0	0	0	0	0	0
S7	1	1	0	1	1	1	0	0	1	1	1	1
S8	1	1	1	1	1	1	1	0	1	1	1	1
S9	1	1	1	1	1	1	0	0	0	1	1	1
S10	0	0	0	0	0	1	0	0	0	0	0	0
S11	1	1	1	1	1	1	0	0	1	1	1	1
S12	0	1	1	0	1	1	0	0	0	1	0	1

Table 13 The final ranking of the substitutes for the crop residues burning

Substitutes	Non-dominant service providers	Rank
S1	S3, S5, S6	**7**th
S2	S5, S6	8th
S3	S6	10th
S4	S1, S2, S3, S5, S6, S10	5th
S5	S3, S6	9th
S6	–	Outranked
S7	S1, S2, S4, S5, S6, S9, S10, S11, S12	2nd
S8	S1, S2, S3, S4, S5, S6, S7, S9, S10, S11, S12	1st
S9	S1, S2, S3, S4, S5, S6, S10, S11	4th
S10	S6	10th
S11	S1, S2, S3, S4, S5, S6, S9, S10, S12	3rd
S1	S2, S3, S5, S6, S10	6th

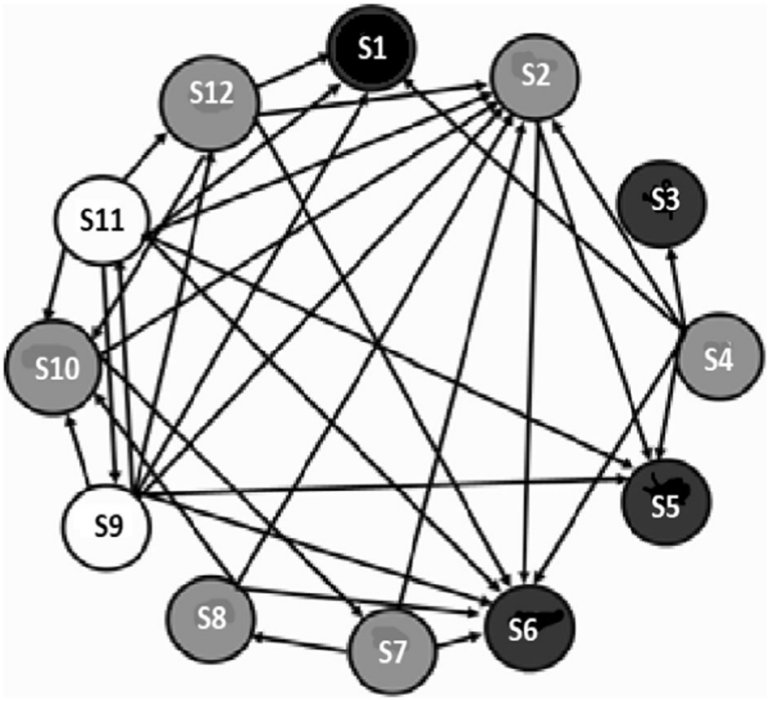


Fig. 3 ELECTRE outranking diagram of the alternatives to burning the crop residues

If $\overline{d}$ is less than d_{kl}, the value in the Boolean matrix will be 1 otherwise 0. Table 11 shows the Boolean matrix F.

The elements of the Boolean matrix E are multiplied with the corresponding elements of the Boolean matrix F (Eq. 12) to prepare a Dominance matrix G. Table 12 shows the dominance matrix G.

On the basis of the dominance matrix G, the ranks are provided to the different Substitute for the crop residues burning as shown in Table 12. The Boolean value 1 indicates the outranking, for example, in row 2, S2 outranks only S5 and S6. It means that S2 is better than the S5 and S6 (Table 13).

Using the outranking an ELECTRE outranking diagram of the substitutes is prepared which is shown in Fig. 3.

It has been observed that the substitute S8 (biogas, biochar and bioethanol production) is ranked as the first preference and S7 (Fuel for Power Plant) is ranked the second preference among all the considered substitutes. These two options may also be considered as the renewable source of energy. It may reduce the need for fossils-fuels up to some extent. Few power plants based on the rice straw as a fuel have already been installed in Punjab. But these power plants are not sufficient to consume all the rice straw produced in the state. The sufficient production facilities for biofuel are needed and suggested to establish in the northern part of India so that some part of requirement of fossil fuels in the form of gasoline or diesel can be fulfilled by the production of biofuel using agro-waste.

The Substitute S11 (preparation of hardboard for the furniture industry) got the third rank among all the other substitutes. Pellets manufacturing from the paddy straw and husk and straw of the wheat (S9) got the fourth rank. The pallets can easily be stored and used as a fuel for the domestic and industrial purposes. Also, the heating value of the pellet is more

than that of the straw of the rice and wheat due to the compact binding. It may be used as a fuel for the industry as well as for the domestic purpose. Use of this substitute for the crop residue burning in the farm will reduce the pollution level. Substitute S6 (Domestic fuel) has been outranked because it requires large space for the storage. It was observed that most of the opinion given by the farmers in FGD regarding the substitutes for the crop residue burning is the industrial applications only. They are not interested to consume the same as compost fertilizers due to large decomposing time. Also, storage of paddy straw as a fuel for the cooking propose is not advisable due to environmental issues and storage problem of the having low density and high volume straw. The industrial application also creates some employment for the unemployed youth of the society.

Other substitutes for the crop residue burning are also not attractive for the farmers due to the long processing time, high logistics cost, high manual labor cost, etc. Substitute S3 (mushroom farming) is also not an attractive solution due to highly perishable product and untraditional type of farming. In addition, the mushroom farming cannot consume all the crop residues at large scale. Amalgamation in soil/in situ incorporation or back in the soil is the time-consuming process and takes time to be decomposed. Using the straw in plastic industry is also in the testing phase. Although it has been reported that these alternatives are in use in some places in China, further research needs to be to be explored. Thus, some more attention is required to develop some infrastructure that can consume and store all the crop residues produced by the farmers throughout the year.

Recently, Govt. of India has announced a crop residue management scheme of 164 million US$ to incentivise farmers in these northern states of India. It consists of subsidies of 80% in purchasing of machines by cooperative groups and 50% by individual farmers. Although this scheme had been useful for big farmers only, small farmers are still not able to purchase these machines for use of crop residues, as they are not able to afford it even after availing subsidies. Due to unavailability of a cost-effective substitute, most of the small farmers prefer burning of crop residues in their fields in spite of the penalty imposed by the Government. To control this problem of crop residue burning, the government should encourage setting up of start-ups in these areas for complete utilization of crop residues. As farmers want to clean their fields for the next crop at earliest, the process of using crop residue should be cost-effective and fast.

Thus, it is observed that farmers are not interested to reuse the crop residue on large scale as in situ management, surface mulching, back in soil, composting, etc., due to long time requirement for composting. They want to clear the farm as soon as possible after harvesting of the rice. In this situation, the stubble and other crop residues may be used for bioenergy, fuel for power plants, and other industrial purpose.

5 Conclusions

In this paper, the issues of crop residue burning in the state of Punjab, Haryana, Rajasthan, and western Uttar Pradesh, i.e., Delhi NCR, have been highlighted. The authors tried to analyze the conditions responsible for air pollution due to the burning of the crop residues.

It has been observed that the major reasons for burning the crop residues are the lack of infrastructure for the consumption of the paddy straws and stubble, lack of cost-effective solutions, and lack of sensitivity for adverse effects of residues burning. The adverse effect of crop residue burning on the environment and consequently health hazard has been also discussed. Some of the important sustainable uses of the crop residues have been suggested with their pros and cons. The general substitutes for using crop residue burning are an amalgamation of stubble in soil/in situ incorporation, crop stubble as animal fodder/cattle shed and bed preparation, wheat and paddy straw in mushroom farming, rice straw mulching, back in the soil after composting, domestic fuel, fuel for power plants, biogas and biochar/bioethanol preparation, making pellets, raw materials for paper and pulp industry, mixing with Plastics.

A hybrid outranking approach comprising of AHP and ELECTRE has been used to rank the various substitutes for use of the crop residues. The input required for the analysis was taken from the opinion of FGD. It has been observed that preparation of biogas, biochar, and bioethanol is the most appropriate option for using crop residues and second most appropriate action is to use the paddy straw as the fuel for the power plant. These substitutes can also be treated as a renewable source of energy. The government needs to develop the required infrastructure in these regions to explore the sustainable usage of crop residues. The government should incentivize farmers and small entrepreneurs to set up the business for use of crop residues for making biogas and bioethanol. It will also solve the problem of unemployment of the youths in rural areas. Currently, farmers find fine imposed as a cheaper option than the cost of lifting straw. In addition to this, the use of other alternatives is time-consuming. To get a second round of crop in their fields, they prefer burning of crop residues as easiest and cheapest alternative. Burning of crop residues requires minimum time to clear the fields. Therefore, farmers should be motivated for alternative use of crop residues. Incentives and awareness programs could be launched by the state and central governments to sensitize the farmers for sustainable use of crop residues.

5.1 Limitations of the research

The complete research is based on the focused group discussion in which farmers, academicians, and industrial participants are involved. A general survey of the farmers may be used to get the information, and statistical analysis can be done to get the useful inference.

5.2 Future work

A broad participation of the farmers from all levels may be considered for the opinion. Some more industrial feasible applications of the stubble other than residue in the farms can be considered as the alternate solutions for open burning of the residue.

Appendix 1

See Fig. 4.

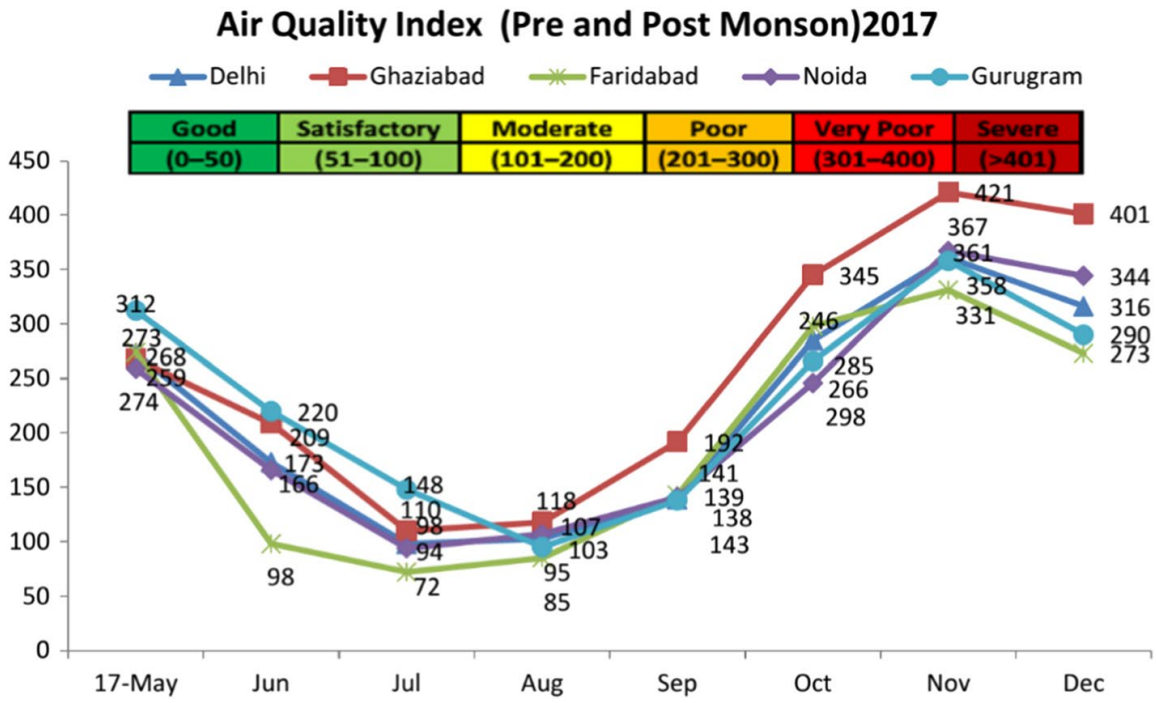


Fig. 4 Pre- and post-monsoon Air Quality Index-2017 in Delhi NCR. *Source*: CPCB (2017)

Appendix 2: Questionnaire

Appendix B: Questionnaire

1. What is your opinion about the worst air-quality of Delhi NCR during the October-November?
2. What are the factors responsible for the degradation of air quality of Delhi NCR?
3. Do you think that stubble burning is one of the main factors for decreasing air- quality of Delhi NCR? Please reply in terms of yes/No.
4. What are the factors compelled the farmers to burn the crop residues (rice stubbles)?
5. What are the alternative options available with you to prevent the stubble burning?
6. If some alternative options are given to you in place of stubble burning, please suggest its suitability and limitations.
7. Please compare the following criteria to evaluate the given alternatives for the stubble burning, on 9-point rating scale as given below.

Importance Scale	Definition of Importance Scale
1	Equally Important Preferred
2	Equally to Moderately Important Preferred
3	Moderately Important Preferred
4	Moderately to Strongly Important Preferred
5	Strongly Important Preferred
6	Strongly to Very Strongly Important Preferred
7	Very Strongly Important Preferred
8	Very Strongly to Extremely Important Preferred
9	Extremely Important Preferred

7.1 Cost of Residue Management Vs. Time of residue composting ☐

7.2 Cost of Residue Management Vs. Air quality degradation ☐

7.3 Cost of Residue Management Vs. Loss of Soil Fertility ☐

7.4 Cost of Residue Management Vs. In-house application ☐

7.5 Cost of Residue Management Vs. Industrial application ☐

7.6 Cost of Residue Management Vs. Need of logistics facility ☐

7.7 Cost of Residue Management Vs. Bio-energy and power production ☐

7.8 Cost of Residue Management Vs. Govt. regulations to prevent stubble burning ☐

7.9 Time of residue composting Vs. Air quality degradation ☐

7.10 Time of residue composting Vs. Loss of soil fertility ☐

7.11 Time of residue composting Vs. In-house application ☐

7.12 Time of residue composting Vs. Industrial application ☐

7.13 Time of residue composting Vs. Need of logistics facility ☐

7.14 Time of residue composting Vs. Bio-energy and power production ☐

7.15 Time of residue composting Vs. Govt. regulations to prevent stubble burning ☐

7.16 Air quality degradation Vs. Loss of soil fertility ☐

7.17 Air quality degradation Vs. In-house application ☐

7.18 Air quality degradation Vs. Industrial Application ☐

7.19 Air quality degradation Vs. Need of logistics facility ☐

7.20 Air quality degradation Vs. Bio-energy and power production ☐

7.21 Air quality degradation Vs. Govt. regulations to prevent stubble burning ☐

7.22 Loss of Soil Fertility Vs. In-house application ☐

7.23 Loss of Soil Fertility Vs. Industrial Application ☐

7.24 Loss of Soil Fertility Vs. Need of logistics facility ☐

7.25 Loss of Soil Fertility Vs. Bio-energy and power production ☐

7.26 Loss of Soil Fertility Vs. Govt. regulations to prevent stubble burning ☐

7.27 In-house application Vs. Industrial Application ☐

7.28 In-house application Vs. Need of logistics facility ☐

7.29 In-house application Vs. Bio-energy and power production ☐

7.30 In-house application Vs. Govt. regulations to prevent stubble burning ☐

7.31 Industrial application Vs. Need of logistics facility ☐

7.32 Industrial application Vs. Bio-energy and power production ☐

7.33 Industrial application Vs. Govt. regulations to prevent stubble burning ☐

7.34 Need of logistics facility Vs. Bio-energy and power production ☐

7.35 Need of logistics facility Vs. Govt. regulations to prevent stubble burning ☐

7.36 Bio-energy and power production Vs. Govt. regulations to prevent the stubble burning ☐

8. Please rate the following alternative solutions for stubble burning with respect to the cost of residue management on 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. And bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

9. Please rate the following alternative solutions for stubble burning with respect to the time of residue Composting/ processing on 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5

(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

10. Please rate the following alternative solutions for stubble burning with respect to the air quality degradation on 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

11. Please rate the following alternative solutions for stubble burning with respect to the loss of soil fertility on 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

12. Please rate the following alternative solutions for stubble burning with respect to In-house application 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5

(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

13. Please rate the following alternative solutions for stubble burning with respect to Industrial applications 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

14. Please rate the following alternative solutions for stubble burning with respect to the need of logistics facilities 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

14. Please rate the following alternative solutions for stubble burning with respect to the bio-energy and power production 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5

(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

14. Please rate the following alternative solutions for stubble burning with respect to the Govt. Regulations to straw burning 5-point rating scale as 1 is least favorable and 5 is most favorable rating. Mark (√) on appropriate rating.

(a) Amalgamation of stubble in Soil/In situ incorporation	1	2	3	4	5
(b) Animal Fodder/shed/bed preparation	1	2	3	4	5
(c) Mushroom Farming	1	2	3	4	5
(d) Rice straw mulching	1	2	3	4	5
(e) Back in the soil after composting	1	2	3	4	5
(f) Domestic fuel	1	2	3	4	5
(g) Fuel for the power plant	1	2	3	4	5
(h) Biogas, biochar. and bioethanol preparation	1	2	3	4	5
(i) Making Pellets	1	2	3	4	5
(j) Mixing with plastics	1	2	3	4	5
(k) Preparation of hardboard for furniture industry	1	2	3	4	5
(l) Paper and pulp industry	1	2	3	4	5

References

Amador, J. A., Glucksman, A. M., Lyons, J. B., & Gorres, J. H. (1997). *Soil Science, 162*(11), 808. https://doi.org/10.1097/00010694-199711000-000057.

Arora, V. K., Singh, C. B., Sidhu, A. S., & Thind, S. S. (2011). Irrigation, tillage and mulching effects on soybean yield and water productivity in relation to soil texture. *Agricultural Water Management, 98*, 563–568.

Beri, V., Sidhu, B. S., Bhat, A. K., &Singh, B. P. (1992). Nutrient balance and soil properties as affected by management of crop residues. In M. S. Bajwa et al. (Eds.), *Nutrient management for sustained productivity. Proceedings of international symposium* (Vol. II, pp. 133–135). Ludhiana: Department of Soil, Punjab Agricultural University.

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus groups in social research*. New Delhi: Sage.

Brar, S. S., Kumar, S., & Narang, R. S. (2000). Effect of moisture regime and nitrogen on decomposition of combine harvested rice residue and performance of succeeding wheat in rice–wheat system in Punjab. *Indian J Agronomy, 45*, 458–462.

Brechbill, S. C., & Tyner, W. E. (2008). *The economics of biomass collection, transportation and supply to Indiana cellulosic and electric utility facility*. Working Paper #08-03. Department of Agricultural Economics, Purdue University.

Campbell, K. A. (2007). *Feasibility study guide for an agricultural biomass Pellet Company*. Waseca: Agricultural Utilization Research Institute.

Chauhan, S. (2011). Biomass resources assessment for power generation: A case study from Haryana State, India. *Biomass and Bioenergy, 34*(9), 1300–1308.

Chauhan, S. (2012). District wise agriculture biomass resource assessment for power generation: A case study from an Indian state, Punjab. *Biomass and Bioenergy, 37*, 205–212.

Cheewaphongphan, P., Junpen, A., Kamnoet, O., & Garivait, S. (2018). Study on the potential of rice straws as a supplementary fuel in very small power plants in Thailand. *Energies, 11*(2), 270.

Choudhary, M., Dhanda, S., Kapoor, S., & Soni, G. (2009). Lignocellulolytic enzyme activities and substrate degradation by Volvariellavolvacea, The paddy straw mushroom/Chinese mushroom. *Indian J Agricultural Research, 43*(3), 223–226.

CPCB. (December, 2015). *Air pollution in Delhi: An analytical study*. Retrieved January 10, 2018 from https://www.cpcbenvis.nic.in/envis_newsletter/Air%20Quality%20of%20Delhi.pdf.

CPCB. (2016). *Air pollution in Delhi: An analysis*. Retrieved November 17th, 2019 from http://cpcbenvis.nic.in/envis_newsletter/Air%20pollution%20in%20Delhi.pdf.

CPCB. (December, 2017). *Ambient air quality data of Delhi-NCR*. Retrieved January 10, 2018 from https://cpcb.nic.in/openpdffile.php?id=TGF0ZXN0RmlsZS8xODNfMTUxNTU4ODM1OF9tZWRpYXBob3RvODM5LnBkZg==).

DAC. (Nov. 2014). *National policy for management of crop residue*. Retrieved Nov 18th, 2019 from https://agricoop.nic.in/sites/default/files/NPMCR_1.pdf.

DAC. (May, 2019). *Report of the committee on review of the scheme "promotion of agricultural mechanization for in-situ management of crop residue in states of Punjab, Haryana, Uttar Pradesh and NCT of Delhi*. Retrieved November 18th, 2019 fromhttps://farmech.dac.gov.in/revised/1.1.2019/REPORT%20OF%20THE%20COMMITTEE-FINAL(CORRECTED).pdf.

Devi, S., Gupta, C., Jat, S. L., & Parmar, M. S. (2017). Crop residue recycling for economic and environmental sustainability: The case of India. *Open Agriculture., 2*(1), 486–494.

Dhaliwal, H. S., Singh, R. P., & Kaur, H. (2011). Financial assessment of happy Seeder. *Conservation Agriculture News Letters, 17*, 4–5.

Douglas, C. L., & Rickman, R. W. (1992). Estimating crop residue decomposition from air temperature, initial nitrogen content, and residue placement. *Soil Science Society of America Journal, 56*(1), 272–278.

Gadde, B., Bonnet, S., Menke, C., & Garivait, S. (2009). Air pollutant emissions from rice straw open field burning in India, Thailand and the Philippines. *Environmental Pollution, 157*, 1554–1558.

Galanter, M., Levy, H., & Carmichael, G. R. (2000). Impacts of biomass burning on tropospheric CO, NOx, and O_3. *Journal of Geophysical Research, 105*(D5), 6633–6653.

Ghude, S. D., Chate, D. M., Jena, C., Beig, G., Kumar, R., Barth, M. C., et al. (2016). Premature mortality in India due to PM2.5 and ozone exposure. *Geophysical Research Letters, 43*, 4650–4658.

Gupta, P. K., Sahai, S., Singh, N., Dixit, C. K., Singh, D. P., & Sharma, C. (2004). Residue burning in rice–wheat cropping system: Causes and implications. *Current Science India, 87*(12), 1713–1715.

Gupta, R. (2011). *Causes of emissions from agricultural residue burning in North-West India; Evaluation of a technology policy response*. Kathmandu: South Asian Network for Development and Environmental Economics (SANDEE).

Hashemi, S. S., Hajiagha, S. H., & Amiri, M. (2014). Decision making with unknown data: Development of ELECTRE method based on black numbers. *Informatics, 25*(1), 21–26.

Hays, M. D., Fine, P. M., Geron, C. D., Kleeman, M. J., & Gullett, B. K. (2005). Open burning of agricultural biomass; physical and chemical properties of particle-phase emissions. *Atmospheric Environment, 39*, 6747–6764.

Hindustan. (15 Nov., 2019). *Vahano ke dhuen se sabse adhik pradushan (Contribution of vehicle in pollution is the highest)*, p. 02.

IARI. (2012). *Crop residues management with conservation agriculture: potential, constraints and policy needs* (pp. 7–32). New Delhi: Indian Agricultural Research Institute.

Ioannidou, O., & Zabaniotou, A. (2007). Agricultural residues as precursors for activated carbon production—A review. *Renewable and Sustainable Energy Reviews, 11*(9), 1966–2005.

Jain, A. K. (2016). Residue crop (Paddy Straw) burning shrouds NCR. In *Proceedings of the 2nd international seminar on utilizationof non-conventional energy sources for sustainable development of rural areas, ISNCESR*. Parthivi College of Engineering &Management, C.S.V.T. University, Bhilai, Chhattisgarh, India. 16, 17th and 18th March 2016.

Jain, N., Pathak, H., & Bhatia, A. (2014). Sustainable management of crop residues in India. *Current Advances in Agricultural Sciences, 6*, 1–9.

Jat, M. L., Gathala, M. K., Ladha, J. K., Saharawat, Y. S., Jat, A. S., Kumar, V., et al. (2009). Evaluation of precision landlevelingand double zero-till systems in the rice–wheat rotation: Water use, productivity, profitability and soil physical properties. *Soil Research, 105*, 112–121.

Jat, M. L., Kamboj, B. R., Sidhu, H. S., Singh, M., Bana, A., Bishnoi, D. K., et al. (2013). *Operational manual for turbo happy Seeder—Technology for managing crop residues with environmental stewardship*. Mexico: CIMMYT.

Jat, M. L., Saharawat, Y. S., & Gupta, R. (2011). Conservation agriculture in cereal systems of South Asia: Nutrient management perspectives, Karnataka. *The Journal of Agricultural Science, 24*, 100–105.

Jeffery, S., Verheijena, F. G. A., van der Veldea, M., & Bastos, A. C. (2011). A quantitative review of the effects of biochar application to soils on crop productivity using meta-analysis. *Agriculture, Ecosystems & Environment, 144*, 175–187.

Karkania, V., Fanara, E., & Zabaniotou, A. (2012). Review of sustainable biomass pellets production—A study for agricultural residues pellets' market in Greece. *Renewable and Sustainable Energy Reviews, 16*(3), 1426–1436.

Kaur, A. (2017). Crop residue in Punjab agriculture-status and constraints. *J Krishi Vigyan, 5*(2), 22–26. https://doi.org/10.5958/2349-4433.2017.00005.8.

Kaur, D., Bhardwaj, N. K., & Lohchab, R. K. (2017). Prospects of rice straw as a raw material for paper making. *Waste Management, 60*, 127–139.

Koopman, A., & Koppejan, J. (1997). *Agricultural and forest residues generation, utilization and availability*. Paper presented at the regional consultation on modern applications of biomass energy, Kuala Lumpur, Malaysia.

Kumar, A., Cameron, J. B., & Flynn, P. C. (2003). Biomass power cost and optimum plant size in Western Canada. *Biomass and Bioenergy, 24*(6), 445–464.

Kumar, K., & Goh, K. M. (1999). Crop residues and management practices: Effects on soil quality, soil nitrogen dynamics, crop yield, and nitrogen recovery. *Advances in Agronomy, 68*, 197–319.

Kumar, P., Kumar, S., & Joshi, L. (2015a). *Socioeconomic and environmental implications of agricultural residue burning: A case study of Punjab, India*, p. 144. Springer Open.

Kumar, P., Singh, R. K., & Kharab, K. (2017a). A comparative analysis of operational performance of cellular mobile telephone service providers in Delhi working area using an approach of fuzzy ELECTRE. *Applied Soft Computing, 59*, 438–447.

Kumar, P., Singh, R. K., & Sinha, P. (2016). Optimal site selection for a hospital using a fuzzy extended ELECTRE approach. *Journal of Management Analytics, 3*(2), 115–135.

Kumar, P., Singh, R. K., & Vaish, A. (2017b). Suppliers' green performance evaluationusing fuzzy extended ELECTRE approach. *Clean Technologies and Environmental Policy, 19*(3), 809–8021.

Kumar, V., Saharawat, Y. S., Gathala, M. K., Jat, A. S., Singha, S. K., Chaudhary, N., et al. (2013). Effect of different tillage and seeding methods on energy use efficiency and productivity of wheat in the Indo-Gangetic Plains. *Field Crops Research, 142*, 1–8.

Lefroy, R. D., Chaitep, W., & Blair, G. J. (1994). Release of sulphur from rice residue under flooded and non-flooded soil conditions. *Australian Journal of Agricultural Research, 45*, 657–667.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature, 525*, 367–371.

Liu, E. (2008). *Straw procurement business case*. Manitoba Bioproducts Working Group.

Liu, W., Xu, Z., & Yang, T. (2018). Health effects of air pollution in China. *International Journal of Environmental Research and Public Health, 15*(7), 1471.

Lohan, S. K., Dixit, J., Kumar, R., Pandey, Y., Khan, J., Ishaq, M., et al. (2015a). Biogas: A boon for sustainable energy development in India's cold climate. *Renewable and Sustainable Energy Reviews, 43*, 95–101.

Lohan, S. K., Dixit, J., Modasir, S., & Ishaq, M. (2012). Resource potential and scope of utilization of renewable energy in Jammu and Kashmir, India. *Renewable Energy, 39*, 24–29.

Lohan, S. K., Jat, H. S., Yadav, A. K., Sindhu, H. S., Jat, M. L., Chaudhary, M., et al. (2018). Burning issues of paddy residue management in north-west states of India. *Renewable and Sustainable Energy Reviews, 81*, 693–706.

Lohan, S. K., Narang, M. K., Manes, G. S., & Grover, N. (2015b). Farmpower availability for sustainable agriculture development in Punjab state of India. *Agricultural Engineering International CIGRJ, 17*(3), 196–207.

Long, W., Tate, R., Neuman, M., Manfreda, J., Becker, A., & Anthonisen, N. (1998). Respiratory symptoms in a susceptible population due to burning of agricultural residue. *Chest, 113*(2), 351.

Mishra, A. K., & Shibata, T. (2012). Synergistic analyses of optical and microphysical properties of agricultural crop residue burning aerosols over the Indo-Gangetic Basin (IGB). *Atmospheric Environment, 57*, 205–218.

Mittal, S. K., Singh, N., Agarwal, N., Awasthi, A., & Gupta, P. K. (2009). Ambient air quality during wheat and rice crop stubble burning episodes in Patiala. *Atmospheric Environment, 43*(2), 238–244.

MNRE. (2009). *Ministry of new and renewable energy resources*. New Delhi: Govt. of India. Retrieved from www.mnre.gov.in/biomassrsources.

NAAS. (2012). *Management of crop residues in the context of conservation agriculture*. Policy Paper No. 58, National Academy of Agricultural Sciences, New Delhi 12.

NDTV. (March 28, 2016). Penalty imposed. In *1406 cases of stubble burning in Haryana*. Accessed March 28, 2016 from https://www.ndtv.com/india-news/penalty-imposed-in-1-406-cases-of-stubble-burning-in-haryana-1621766.

Oanh, N. T. K., Ly, B. T., Tipayarom, D., Manandhar, B. R., Prapat, P., Simpson, C. D., et al. (2011). Characterization of particulate matter emission from open burning of paddy straw. *Atmospheric Environment, 45*, 493–502.

Oanh, N. T. K., Permadi, D. A., Hopke, P. K., Smith, K. R., Dong, N. P., & Dang, A. N. (2018). Annual emissions of air toxics emitted from crop residue open burning in Southeast Asia over the period of 2010–2015. *Atmospheric Environment, 187*, 163–173.

Pandey, C. (2019). Management of crop residue for sustaining soil fertility and foodgrains production in India. *Acta Scientific Agriculture, 3*(1), 188–195.

Pandey, C. H., & Sujatha, D. (2011). *Crop residues, the alternate raw materials of tomorrow for the preparation of composite board*. Indian Plywood Industries Research & Training Institute. https://innovate.mygov.in/wp-content/uploads/2018/07/mygov1532797148376098.pdf

Pathak, H., Bhatia, A., & Jain, N. (2010). *Inventory of greenhouse gas emission from agriculture*. Report submitted to Ministry of Environment and Forests, Govt. of India.

Pathak, H., Singh, R., Bhatia, A., & Jain, N. (2006). Recycling of rice straw to improve wheat yield and soil fertility and reduce atmospheric pollution. *Paddy and Water Environment, 4*(2), 111.

PSCST-TERI. (2014). *Action plan for green budgeting in Punjab; concepts, rationale and ways forward*. The Energy and Resources Institute (TERI) and Punjab State Council for Science and Technology (PSCST). Supported by Department of Science, Technology and Environment, Government of Punjab, India.

Ramanathan, V., & Carmichael, G. (2008). Global and regional climate changes due to black carbon. *Nature Geo Science, 1*(4), 221–227.

Ravindra, K., Singh, T., & Mor, S. (2018). Emissions of air pollutants from primary crop residue burning in India and their mitigation strategies for cleaner emissions. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2018.10.031.

Roy, P., & Kaur, M. (2016). Assessment of farmers' knowledge about beneficial effects of application of paddy straw onto soil in West Bengal. *Journal of Community Mobilization and Sustainable Development, 11*(1), 100–106.

Saaty, T. L. (1980). *The analytical hierarchy process, planning, priority. Resource allocation*. Pittsburgh: RWS Publications.

SAFAR. (2019). *System of air quality and weather forecasting and research*. Retrieved November 17th, 2019 from https://safar.tropmet.res.in/.

Sahai, S., Sharma, C., Singh, D. P., Dixit, C. K., Singh, N., Sharma, P., et al. (2007). A study for development of emission factors for trace gases and carbonaceous particulate species from in situ burning of wheat straw in agricultural fields in India. *Atmospheric Environment, 41*, 9173–9186.

Sangeet, K. R. (2016). Crop residue generation and management in Punjab state. *Indian Journal of Economics and Development, 12*(1), 477–483.

Sekhon, N. K., Singh, C. B., Sidhu, A. S., Thind, S. S., Hira, G. S., & Khurana, D. S. (2008). Effect of mulching, irrigation and fertilizer nitrogen levels on soil hydrothermal regime, water use and yield of hybrid chilli. *Archives of Agronomy and Soil Science, 54*, 163–174.

Shanian, A., & Savadogo, O. (2006). ELECTRE I decision support model for material selection of bipolar plates for polymer electrolyte fuel cells applications. *Journal of New Materials for Electrochemical Systems, 9*, 191–199.

Sidhu, B. S., & Beri, V. (2005). Experience with managing rice residues in intensive rice–wheat cropping system in Punjab. In I. P. Abrol, R. K. Gupta, & R. K. Malik (Eds.), *Conservation agriculture: Status and prospects* (pp. 55–63). New Delhi: Centre for Advancement of Sustainable Agriculture, National Agriculture Science Centre.

Sidhu, H. S., Singh, S., Singh, Y., Blackwell, J., Lohan, S. K., Humphreys, E., et al. (2015). Development and evaluation of the turbo happy Seeder for sowing wheat into heavy rice residues in NIndia. *Field Crops Research, 184*, 201–212.

Singh, A. P., Chakrabarti, S., Kumar, S., & Singh, A. (2017). Assessment of air quality in Haora River basin using fuzzy multiple-attribute decision making techniques. *Environmental Monitoring and Assessment, 189*(8), 373.

Singh, C. P., & Panigrahy, S. (2011). Characterisation of residue burning from agricultural system in India using space based observations. *Journal of the Indian Society of Remote Sensing, 39*(3), 423–429.

Singh, J., Panesar, B. S., & Sharma, S. K. (2008). Energy potential through agricultural biomass using geographical information system—A case study of Punjab. *Biomass and Bioenergy, 32*, 301–307.

Singh, K. K., Lohan, S. K., Jat, A. S., & Rani, T. (2006). New technologies of growing rice for higher production. *Research Crops, 7*(2), 369–371.
Singh, M., Sidhu, H. S., Singh, Y., & Blackwell, J. (2011). Effect of rice straw management on crop yields and soil health in rice–wheat system. *Conservation Agricultural News PACA, 18*, 1–11.
Singh, R. P., & Kaskaoutis, D. G. (2014). Crop residue burning: a threat to South Asian air quality. *EOS, Transactions American Geophysical Union*, *95*(37), 333–340.
Singh, S., Batra, R., Mishra, M. M., Kapoor, K. K., & Goyal, S. (1992). Decomposition of paddy straw in soil and the effect of straw incorporation in the field on the yield of wheat. *Journal of Plant Nutrition and Soil Sciences, 155*(4), 307–311.
Singh, Y., Humphreys, E., Kukal, S. S., Singh, B., Kaur, A., Thaman, S., et al. (2009). Crop performance in permanent raised bed rice–wheat cropping system in Punjab India. *Field Crops Research, 110*, 1–20.
Singh, Y., & Sidhu, H. S. (2014). Management of cereal crop residues for sustainable rice–wheat production system in the Indo-Gangetic plains of India. *Proceeding of Indian National Science Academy, 80*(1), 95–114.
Singh, Y., Sidhu, H. S., Khanna, P. K., Kapoor, S., Jain, A. K., Singh, A. K., et al. (2010). *Options for effective utilization of crop residues*. Ludhiana: Directorate of Research Punjab Agricultural University.
Singh, Y., Singh, B., Ladha, J. K., Khind, C. S., Khera, T. S., & Bueno, C. S. (2004). Management effects on residue decomposition, crop production and soil fertility in a rice–wheat rotation in India. *Soil Science Society of America Journal, 68*, 320–326.
Singh, Y., Singh, B., & Timsina, J. (2005). Crop residue management for nutrient cycling and improving soil productivity in paddy-based cropping systems in the tropics. *Advances in Agronomy, 85*, 269–407.
Singh, Y., Singh, D., & Tripathi, R. P. (1996). Crop residue management in rice–wheat cropping system. In *Abstracts of poster sessions 2nd international crop science congress* (p. 43). New Delhi: National Academy of Agricultural Sciences, p. 26.
Stewart, D. W., & Shamdasani, P. N. (1990). *Focus groups: theory and practices*. London: Sage.
Sultana, A., Kumar, A., & Harfield, D. (2010). Development of agri-pellet production cost and optimum size. *Bioresource Technology, 101*(14), 5609–5621.
Tabatabai, M. A. (1994). Soil enzymes. In A. L. Page, R. H. Miller, & D. R. Keeney (Eds.), *Methods of soil analysis* (pp. 775–833). Madison: American Society of Agronomy.
Teodoro, C., & Mendoza, B. C. (2016). A review of sustainability challenges of biomass for energy: Focus in the Philippines. *Journal of Agricultural Technology, 12*(2), 281–310.
Tianjia, L., Miriam, E., Marlier, R. S., DeFries, D. M., Westervelt, K. R., Xia, A. M., et al. (2018). Seasonal impact of regional outdoor biomass burning on air pollution in three Indian cities: Delhi, Bengaluru, and Pune. *Atmospheric Environment, 172*, 83–92.
Treets, D. G., Yarber, K. F., Woo, J. H., & Carmichael, G. R. (2003). Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions. *Global Biogeochemical Cycles, 17*, 1099.
Vadrevu, K. P., Ellicott, E., & Badarinath, K. (2011). MODIS derived fire characteristics and aerosol optical depth variations during the agricultural residue burning season, North India. *Environmental Pollution, 159*(6), 1560–1569.
Van den Elshout, S., Léger, K., & Heich, H. (2014). CAQI common air quality index—Update with PM2. 5 and sensitivity analysis. *Science of the Total Environment, 488*, 461–468.
Verma, S. S. (2014). Technologies for stubble use. *Journal of Agriculture and Life Sciences, 1*(1), 106–110.
WHO. (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. ISBN 9789241511353. https://apps.who.int/iris/bitstream/handle/10665/250141/9789241511353-eng.pdf?sequence=1
Yang, S., He, H., Lu, S., Chen, D., & Zhu, J. (2008). Quantification of crop residue burning in the field and its influence on ambient air quality in Suqian, China. *Atmospheric Environment, 42*(9), 1961–1969.
Yang, S. S., Liu, C. M., Lai, C. M., & Liu, Y. L. (2003). Estimation of methane and nitrous oxide emission from paddy fields and uplands during 1990–2000 in Taiwan. *Chemosphere, 52*, 1295–1305.
Zhang, H. F., Ye, X. N., Cheng, T. T., Chen, J. M., Yang, X., Wang, L., et al. (2008). A laboratory study of agricultural crop residue combustion in China: Emission factors and emission inventory. *Atmospheric Environment, 42*, 8432–8441.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 6, Issue 1; January-March, 2019, pp. 45-48
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/Publication.html

Solar Thermal Collector Integrated Organic Rankine Cycle Technology

Harwinder Singh[1] and R.S. Mishra[2]

[1]Department of Mechanical Engineering
Delhi Technological University
New Delhi-110042
[2]Department of Mechanical Engineering
Delhi Technological University
New Delhi-110042
E-mail: [1]harrymehrok14@gmail.com, [2]rsmishra@dce.ac.in

Abstract—*Nowadays, organic Rankine cycle (ORC) has emerged as an option for power production and waste heat recovery. This review paper deals with the working and layouts of ORC, and latest potential research work present in the field of ORC. Different configurations types of ORC and application of solar energy as a heat source has been discussed. Further applications of ORC in the combined cycle has been reviewed in this paper.*

1. INTRODUCTION

In recent years, renewable energy sources have attracted more attention by converting renewable energy into useful forms of energy. In this way, there are several advantages of solar energy have been noticed such as it is accessible in numerous regions round the world and it does not cause environmental pollution, therefore, it is known as clean energy [1]. Nowadays as well as in future, in terms of power generation, solar energy can be considered as most promising and viable option [2]. Solar photovoltaic systems and solar thermal systems are two commercial methods used to convert solar energy into power, and in solar thermal device, solar energy either directly utilized in heating application or mechanical work could be generate from heat with the help of a power cycle [3]. The Rankine cycle, Brayton cycle or the Stirling engine can be used to convert solar thermal energy into mechanical power. To generate electricity from solar thermal energy, the rankine cycle is assumed as the most usual and competitive power generation cycle [4,5]. The efficient conversion of low grade thermal energy to generate electricity could be possible by thermodynamic cycles such as ORC, supercritical Rankine Cycle, Kalina cycle, and Trilateral Flash Cycle [6].

In order to recover waste heat, the well-known ORC is a promising solution because of its flexibility, high safety, low maintenance requirements and good thermal performance like advantages [7-9]. The ORC uses refrigerants or volatile organic fluids instead of water as in conventional Rankine cycle [10-12]. Because as compared to water, organic working fluids due to their lower boiling points can possibly recover energy from the waste heat sources with low temperature [13].

In a solar operated ORC, working fluid at high pressure is heated and evaporated by the solar radiations, after which mechanical shaft work is generated by the expansion of vapours. This shaft work is used to drive a pump or used to produce electricity with the help of a generator [14]. In this paper, detailed literature survey on the solar integrated ORC and its different layouts along with their various aspects has been presented.

2. WORKING OF ORC AND ITS LAYOUTS

The condenser, expander, pump, working fluid and heat source are the major components of an ORC [16]. The organic working fluid pressurized by using a pump and sent to the evaporator (process 1-2). The organic working fluid at high pressure after heated by a heat source, convert into high pressure and high temperature vapours enters in the turbine (process 2-3). The kinetic energy of the fluid used to rotate the turbine blades, then converted to mechanical energy which is further used to operate the generator in order to produce power (process 3-4). An external source is used to cooled down the exhaust from the turbine outlet to bring organic fluid to initial condition, as a result, cycle completes (process 4-1) [6]. The different layouts of ORC such as subcritical and transcritical cycles are shown in Fig. 1 and SORC is shown in Fig. 2.

Other types of ORC are Single stage regenerative ORC (SSRORC), Double stage regenerative ORC (DSRORC), Reheat ORC (RORC), and dual loop ORC (DLORC) [17].

In SSRORC system, some vapours are taken out between two stages of the turbine and added into the feed water heater [17-20] as shown in Fig. 3. The cycle efficiency can be enhanced by reducing the heat addition from the evaporator heat source

with the help of a regenerator [17,21]. The DSRORC system is same as SSRORC system, only difference is that extraction will be happened between two stages as illustrated in Fig. 4. This configuration can enhance the efficiency of cycle by decreasing the evaporator load [17,19-21].

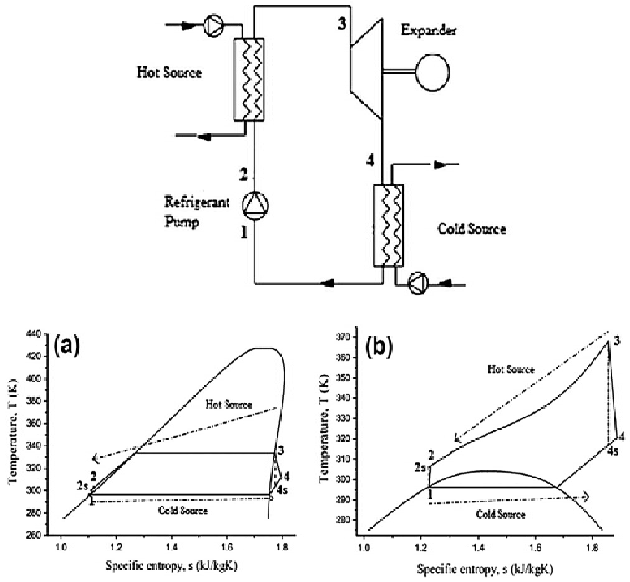


Figure 1: Layout of ORC and temperature-entropy diagrams (a) subcritical cycle (b) transcritical cycle [6,15]

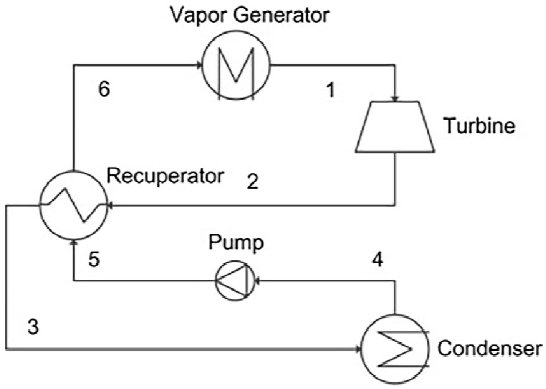


Figure 2: Diagram of SORC system [6]

In RORC, vapours at high pressure from the evaporator enters into the first turbine. After this, the outlet vapours again enter into the evaporator and where these vapours undergo reheated process by using the heat source before arriving to the second turbine with lower pressure as shown in Fig. 5. The RORC system has a purpose to eliminate the steam's moisture at the expansion process' final stages [17,21]. The high temperature loop in DLORC has been utilized to recover the waste heat source as illustrated in Fig. 6. To recover the cooling water of jacket and the HT loop's excess heat, the LT loop can be utilized. This system enhances the cycle's overall efficiency by reducing the heat load that dissipated to the environment [17,22-24].

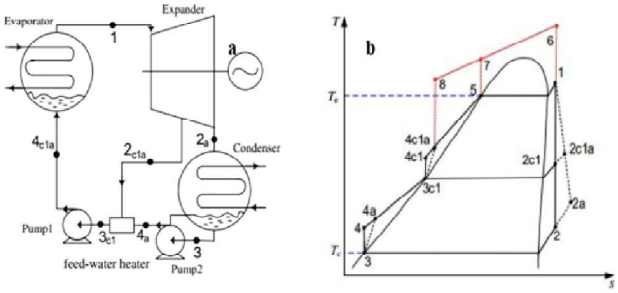


Figure 3: (a) Diagram of SSRORC (b) temperature-entropy diagram for SSRORC [17,19]

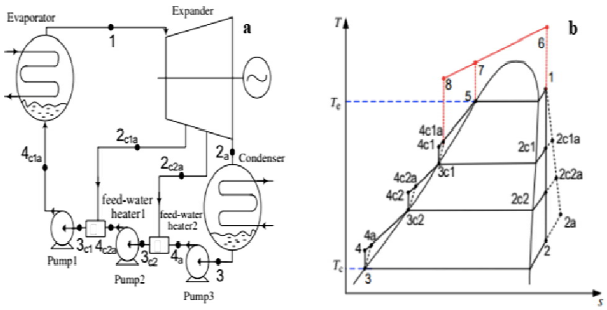


Figure 4: (a) Diagram of DSRORC (b) temperature-entropy diagram for DSRORC [17,19]

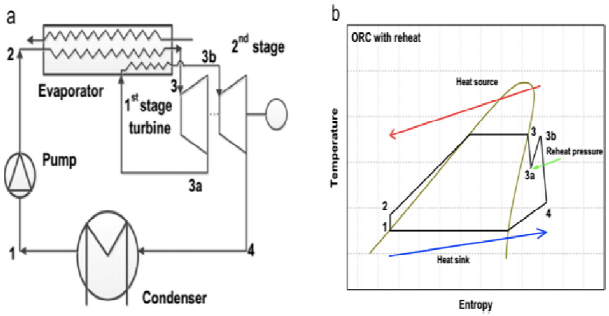


Figure 5: (a) Diagram of RORC (b) temperature-entropy diagram for RORC [17,21]

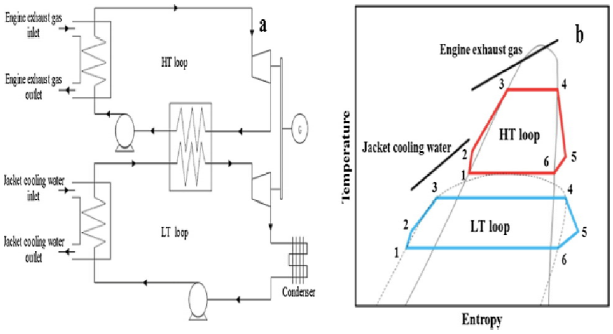


Figure 6: (a) Diagram of DLORC (b) temperature-entropy diagram for DLORC [17,24]

3. SOLAR ENERGY BASED ORC SYSTEM

Singh and Mishra [25] conducted an exergy and energy based analysis of parabolic trough collector integrated supercritical ORC. They found that R600a based supercritical ORC has the maximum exergy efficiency around 96.09% at solar direct normal irradiance of 0.95 kW/m^2. Ramos et al. [26] investigated the solar combined heat and power system based on ORC engine and found that by combining an evacuated-tube collector array and an ORC engine can deliver an electrical output of 3,605 kWh/year from collector array of 60 m^2. He et al. [27] proposed a cogeneration system with a single-extraction ORC as the top cycle and found that the cogeneration system's net power, production of water, and total efficiency found to be as 15.74 kW, 76.88 kg/h and 40.81%, respectively. Morrone et al. [28] analysed the performance of a micro-hybrid energy system for combined heat and power generation, and apparatus is based on a transcritical ORC which is fuelled by a conventional biomass boiler and parabolic trough collectors. They found that hybridisation of biomass and solar allows reducing the consumption of biomass, increasing the maximum operating hours, and improving the efficiency of global system. Freeman et al. [29] examined the domestic-scale solar combined heat and power system based on an ORC engine. They found that with the implication of evacuated flat-plate collectors, the overall solar-to-electrical conversion efficiencies found to be in the range 4.4–6.4% for UK and 6.3–7.3% for Cyprus. Bryszewska-Mazurek et al. [30] performed an analysis of a power plant with the ORC and found that ORC with a heat regenerator has the maximum thermodynamic efficiency of ~ 9%.

4. ORC APPLICATION IN COMBINED CYCLE

Singh and Mishra [31] performed an energy and exergy analysis of solar parabolic trough collectors driven combined supercritical CO_2 cycle and ORC as a bottoming system. They found that R407c based combined system has the highest exergy and thermal efficiency which is 78.07% and 43.49% at solar direct normal irradiance of 0.95 kW/m^2, respectively. In a different study, Singh and Mishra [32] replaced the simple configuration by recompression cycle in the combined system and found that R123 based combined recompression SCO_2 cycle and ORC system demonstrates the highest thermal and exergy efficiency which is around 73.4% and 40.89% at direct normal irradiance of 0.5 kW/m^2. Njoku et al. [33] performed an analysis of a combined gas- and steam- turbine cycle power plant integrated with ORC and absorption refrigeration cycle. They found by utilizing exhaust heat of the combined gas- and steam- turbine cycle power plant to power an R113 fluid based ORC, extra electricity of 7.5 MW was generated. Liu et al. [34] investigated the thermodynamic performance of ORC power generation system which has been used to recover and utilize the waste heat of the intercooled system and found that maximum enhancements of output power and thermal efficiency were 6.08% and 2.14%, respectively. Al-Sulaiman [35] conducted an exergy analysis of parabolic trough collectors integrated combined steam Rankine cycle and ORC. They found that the best exergy performance was possessed by the R134a based combined cycle with maximum value of exergy efficiency, i.e. 26% followed by the combined cycle based on R152a fluid with 25%, and R600a based combined cycle has the lowest exergy efficiency, i.e. 20-21%. In a different study, Al-Sulaiman [36] conducted an energy and solar field sizing analysis and found that due to best performance of R134a combined cycle, it requires the smallest size of solar field.

5. CONCLUSION

In this paper, working of ORC and its various layouts have been discussed. Furthermore, different studies related to ORC powered solar energy has been reviewed from Ref. [25-30]. Moreover, applications of ORC in combined cycle for power generation and waste heat recovery has been reviewed in Ref. [31-36].

REFERENCES

[1] Kalogirou, S. A., "Solar thermal collectors and applications", *Prgress in Energy and Combustion Science*, 30, 2004, pp. 231-295.

[2] Siva Reddy, V., Kaushik, S. C., Ranjan, K. R., Tyagi, S. K., "State-of-the-art of solar thermal power plants - a review", *Renewable and Sustainable Energy Reviews*, 27, 2013, pp. 258–273.

[3] Twomey, B., Jacobs, P. A., Gurgenci, H., "Dynamic performance estimation of small-scale solar cogeneration with an organic Rankine cycle using a scroll expander", *Applied Thermal Engineering*, 51, 2013, pp. 1307–1316.

[4] Tzivanidis, C., Bellos, E., Antonopoulos, K. A., "Energetic and financial investigation of a stand-alone solar-thermal Organic Rankine Cycle power plant", *Energy Conversion and Management*, 126, 2016, pp. 421–433.

[5] Aboelwafa, O., Fateena, S-E. K., Soliman, A., Ismail, I. M., A review on solar Rankine cycles: Working fluids, applications, and cycle modifications", *Renewable and Sustainable Energy Reviews*, 82, 2018, 868–885.

[6] Pethurajan, V., Sivan, S., Joy, G. C., "Issues, comparisons, turbine selections and applications – An overview in organic Rankine cycle", Energy Conversion and Management, 166, 2018, 474–488.

[7] Tchanche, B. F., Lambrinos, G., Frangoudakis, A., Papadakis, G., "Low-grade heat conversion into power using organic Rankine cycles–a review of various applications", *Renewable and Sustainable Energy Reviews*, 15, 2011, 3963–3979.

[8] Tian, H., Shu, G., Wei, H., Liang, X., Liu, L., "Fluids and parameters optimization for the organic Rankine cycles (ORCs) used in exhaust heat recovery of Internal Combustion Engine (ICE)", *Energy*, 47, 2012, pp. 125–136.

[9] Wei, D., Lu, X., Lu, Z., Gu, J., "Performance analysis and optimization of organic Rankine cycle (ORC) for waste heat recovery", *Energy Conversion and Management*, 48, 2007, pp. 1113–1119.

[10] Badr, O., Probert, S., O'callaghan, P., "Selecting a working fluid for a Rankine-cycle engine", *Applied Energy*, 21, 1985, pp. 1–42.
[11] Papadopoulos, A. I., Stijepovic, M., Linke, P., "On the systematic design and selection of optimal working fluids for Organic Rankine Cycles", *Applied Thermal Engineering*, 30, 2010, 760–769.
[12] Vélez, F., Segovia, J. J., Martín, M. C., Antolín, G., Chejne, F., Quijano, A., "A technical, economical and market review of organic Rankine cycles for the conversion of low-grade heat for power generation", *Renewable and Sustainable Energy Reviews*, 16, 2012, pp. 4175–4189.
[13] Nguyen, T., Slawnwhite, J., Boulama, K. G., "Power generation from residual industrial heat", *Energy Conversion and Management*, 51, 2010, pp. 2220–2229.
[14] Freeman, J., Hellgardt, K., Markides, C. N., "An assessment of solar-powered organic Rankine cycle systems for combined heating and power in UK domestic applications", *Applied Energy*, 138, 2015, pp. 605–620
[15] Davide, Z., Asfaw, B., Mauro, V., "Advances and challenges in ORC systems modeling for low grade thermal energy recovery", *Applied Energy*, 121, 2014, pp. 79–95.
[16] Chen, Y., Lundqvist, P., Johansson, A., Platell, P., "A comparative study of the carbon dioxide transcritical power cycle compared with an organic Rankine cycle with R123 as working fluid in waste heat recovery", Applied Thermal Engineering, 26, 2006, pp. 2142–2147.
[17] Mahmoudi, A., Fazli, M., Morad, M. R., "A recent review of waste heat recovery by Organic Rankine Cycle", *Applied Thermal Engineering*, 143, 2018, pp. 660–675.
[18]
[19] Peris, B., Navarro-Esbrí, J., Molés, F., "Bottoming organic Rankine cycle configurations to increase Internal Combustion Engines power output from cooling water waste heat recovery", *Applied Thermal Engineering*, 61, 2013, 364–371.
[20] Xi, H., Li, M-J., Xu, C., He, Y-L, "Parametric optimization of regenerative organic Rankine cycle (ORC) for low grade waste heat recovery using genetic algorithm", *Energy*, 58, 2013, 473–482.
[21] Imran, M., Park, B. S., Kim, H. J., Lee, D. H., Usman, M., Heo, M., "Thermo-economic optimization of Regenerative Organic Rankine Cycle for waste heat recovery applications", *Energy Conversion and Management*, 87, 2014, pp. 107–118.
[22] Li, G., "Organic Rankine cycle performance evaluation and thermoeconomic assessment with various applications part I: energy and exergy performance evaluation", Renewable and Sustainable Energy Reviews, 53, 2016, pp. 477–499.
[23] Song, J., Gu, C-W., "Parametric analysis of a dual loop Organic Rankine Cycle (ORC) system for engine waste heat recovery", *Energy Conversion and Management*, 105, 2015, pp. 995–1005.
[24] Yang, F., Dong, X., Zhang, H., Wang, Z., Yang, K., Zhang, J., Wang, E., Liu, H., Zhao, G., "Performance analysis of waste heat recovery with a dual loop organic Rankine cycle (ORC) system for diesel engine under various operating conditions", *Energy Conversion and Management*, 80, 2014, pp. 243–255.
[25] Zhou, Y., Wu, Y., Li, F., Yu, L., "Performance analysis of zeotropic mixtures for the dual-loop system combined with internal combustion engine", *Energy Conversion and Management*, 118, 2016, pp. 406–414.
[26] Singh, H., Mishra, R. S., "Performance Evaluation of the Supercritical Organic Rankine Cycle (SORC) Integrated with Large Scale Solar Parabolic Trough Collector (SPTC) System: An Exergy Energy Analysis", *Environmental Progress & Sustainable Energy*, 37, 2018, 891-899.
[27] Ramos, A., Chatzopoulou, M. A., Freeman, J., Markides, C. N., "Optimisation of a high-efficiency solar-driven organic Rankine cycle for applications in the built environment", *Applied Energy*, 228, 2018, 755–765.
[28] He, W. F., Wu, F., Kong, Y. P., Wenc, T., Chena, J. J., Han, D., "Parametric analysis of a power-water cogeneration system based on single-extraction organic Rankine cycle", *Applied Thermal Engineering*, 148, 2019, pp. 382–390.
[29] Morrone, P., Algieri, A., Castiglione, T., "Hybridisation of biomass and concentrated solar power systems in transcritical organic Rankine cycles: A micro combined heat and power application", *Energy Conversion and Management* 180, 2019, pp. 757–768.
[30] Freeman, J., Guarracino, I., Kalogirou, S. A., Markides, C. N., "A small-scale solar organic Rankine cycle combined heat and power system with integrated thermal energy storage", *Applied Thermal Engineering*, 127, 2017, pp. 1543–1554.
[31] Bryszewska-Mazurek, A., Świeboda, T., Mazurek, W., "Performance Analysis of a Solar-Powered Organic Rankine Cycle Engine", *Journal of the Air & Waste Management Association*, 61, 2011, pp. 3– 6.
[32] Singh, H., Mishra, R. S., "Performance analysis of solar parabolic trough collectors driven combined supercritical CO2 and organic Rankine cycle", *Engineering Science and Technology, an International Journal*, 21, 2018, pp. 451–464.
[33] Singh, H., Mishra, R. S., "Energy- and exergy-based performance evaluation of solar powered combined cycle (recompression supercritical carbon dioxide cycle/organic Rankine cycle)", *Clean Energy*, 2, 2018, pp. 140-153.
[34] Njoku, I. H., Oko, C. O. C., Ofodu, J. C., "Performance evaluation of a combined cycle power plant integrated with organic Rankine cycle and absorption refrigeration system", *Cogent Engineering*, 5, 2018, pp. 1451426, 1-30.
[35] Liu, W., Zhang, X., Zhao, N., Shu, C., Zhang, S., Ma, Z., Han, J., "Performance analysis of organic Rankine cycle power generation system for intercooled cycle gas turbine", *Advances in Mechanical Engineering*, 10, 2018, pp. 1–12.
[36] Al-Sulaiman, F. A., "Exergy analysis of parabolic trough solar collectors integrated with combined steam and organic Rankine cycles", *Energy Conversion and Management*, 77, 2014, pp. 441–449.
[37] Al-Sulaiman, F. A., "Energy and sizing analyses of parabolic trough solar collector integrated with steam and binary vapor cycles", *Energy*, 58, 2013, pp. 561-570.

arXiv:2004.02219v1 [cs.CL] 5 Apr 2020

Speaker Recognition using SincNet and X-Vector Fusion

Mayank Tripathi, Divyanshu Singh, and Seba Susan (✉)[0000−0002−6709−6591]

Department of Information Technology
Delhi Technological University, Delhi 110042, India
{mayank_bt2k16,divyanshu_bt2k16}@dtu.ac.in, seba_406@yahoo.in

Abstract. In this paper we propose an innovative approach to perform speaker recognition by fusing two recently introduced deep neural networks (DNNs) namely - SincNet and X-Vector. The idea behind using SincNet filters on the raw speech waveform is to extract more distinguishing frequency-related features in the initial convolution layers of the CNN architecture. X-Vectors are used to take advantage of the fact that this embedding is an efficient method to churn out fixed dimension features from variable length speech utterances, something which is challenging in plain CNN techniques, making it efficient both in terms of speed and accuracy. Our approach uses the best of both worlds by combining X- vector in the later layers while using SincNet filters in the initial layers of our deep model. This approach allows the network to learn better embedding and converge quicker. Previous works use either X-Vector or SincNet Filters or some modifications, however we introduce a novel fusion architecture wherein we have combined both the techniques to gather more information about the speech signal hence, giving us better results. Our method focuses on the VoxCeleb1 dataset for speaker recognition, and we have used it for both training and testing purposes.

Keywords: Speaker Recognition · Deep Neural Networks · SincNet · X-vector · VoxCeleb1 · Fusion Model.

1 Introduction

Speaker Recognition and Automatic Speech Recognition(ASR) are two of the actively researched and interesting fields in the computer science domain. Speaker recognition has applications in various fields such as biometric authentication, forensics, security and speech recognition, which has contributed to steady interest in this discipline [1]. The conventional method of speaker identification involves classification of features extracted from speech such as the Mel Frequency Cepstral Coefficients (MFCC) [19].

With the advent of i-vectors [4], speaker verification has become faster and more efficient as compared to the preceding model based on the higher dimensional Gaussian Mixture Model (GMM) supervectors. The i-vectors are low

dimensional vectors that are rich in information and represent distinguishing features of the speaker. i-vectors are used to generate a verification score between the embedding of two speakers. This score gives us information about whether both are the same speaker or different speakers. Previous experiments have shown that i-vectors perform better with Probabilistic Linear Discriminant Analysis (PLDA) [7]. Work has also been carried out in the field of training the i-vectors using different techniques in order to get better embedding, therefore, better results [5,12].

Currently, researchers are moving towards Deep Neural Network (DNN) to obtain speaker embedding. DNN can be directly optimized to distinguish between speakers. DNN showed promising results in comparison to statistical measures such as i-vectors [6,17]. X-vectors are seen as an improvement over the i-vector system (which is also a fixed-dimension vector system) because they are more robust and have yielded better results [18]. X-vector extraction methodology employs a Time-Delayed Neural Network (TDNN) to learn features from the variable-length audio samples and converts them to fixed dimension vectors. This architecture can broadly be broken down in their order of occurrence into three units, namely, frame-level layers, statistics pooling layer and the segment level layers. X- vectors can then be used with classifiers of any kind to carry out recognition tasks.

SincNet [14] is a deep neural network that has embedded band pass filters for extracting features from the audio sample. The features are then fed into DNN based classifiers. We have used SincNet filters and not fed the audio waveform directly into the DNN based classifiers as the latter technique poses problems like high convergence time and less appealing results. The SincNet filters are actually band pass filters which are derived from parameterized sinc functions. This gives us a compact and an efficient way to get a filter bank that can be customized and specifically tuned for a given application.

Our novel architecture combines the goodness of both the methodologies - SincNet and X-Vector, by extracting features using both the techniques and feeding them to fully connected dense layers which acts as a classifier. The organization of this paper is as follows: the related works are described in section 2, the proposed fusion architecture is presented in section 3, the experimental setup and the results are discussed in sections 4 and 5 respectively, and the final conclusions are drawn in section 5.

2 Related Works

The i-vectors [4] feature extraction method has proved to be state-of-the-art for quite some time now for the speaker recognition tasks. Techniques like the Probabilistic Linear Discriminant Analysis (PLDA) and Gauss-PLDA [3,7] are fed with the extracted features and based upon which they carry out classification. Despite being a state-of-the-art technique, we can still improve the performance in terms of accuracy.

With the advent of various deep learning techniques in multiple domains for feature extraction, work has also been carried out to extract features from audio signals using deep learning architectures [9]. These architectures can range from using deep learning just for feature extraction to using neural networks as classifiers as well. The deep learning methods give better results when compared to older techniques based on feature engineering [20].

The most commonly used deep learning method for feature extraction is the one based on the Convolution Neural Network (CNN) architecture [17]. The CNN has been a preferred choice for researchers as it has given good quality results in tasks such as image recognition. Initially the CNN was fed with spectrogram in order to extract features from the audio signals [2,13,15,17,22]. Despite the fact that, spectrogram-based CNN methods were giving good results there were again many drawbacks of using this method [21]. Firstly, the spectrogram is the temporal representation of data unlike images which are spatio-temporal representations. Secondly, a frame in spectrogram can be a result of superposition of multiple waves and not just a single wave which means that a single result can be obtained using different waves super-positioned in different manner. Also, even though spectrograms retain more information than standard hand-crafted features, their design still requires careful tuning of some hyper-parameters, such as the duration, overlap, and typology of the frame window.

The above drawbacks also inspired researchers to directly input raw waveforms into CNN [8] so that no information is lost. This was a good methodology to follow but it results in slower convergence since it processes the complete frequency band of the audio signal.

SincNet [14] and X-vector [18] are amongst the most recent deep learning methods for speech signal classification. Both of them have proved to be more robust than methods which have preceded them.

3 Proposed Fusion Model

We propose a novel fusion of SincNet [14] and X-Vector [18] embedding which will enable us to take the temporal features into account which is quite important for any audio based recognition task since the signal at any point at time t in an audio signal is affected by points preceding and succeeding it.

3.1 X-Vector Embedding

We have used a pre-trained X-vector system which was trained on the VoxCeleb1 [11] dataset which we are using. The pre-trained x-vector system is available in the kaldi toolkit which is available for public use [16]. Table. 1 shows the architecture of the x-vector feature extractor system which has been trained on the VoxCeleb1 dataset. X-vector extraction methodology employs a Time-Delayed Neural Network (TDNN) to learn features from the variable-length audio samples and converts them to fixed dimension vectors. This architecture can broadly be broken down in their order of occurrence into three units, namely, frame-level

layers, statistics pooling layer and the segment level layers. X- vectors can then be used with classifiers of any kind to carry out recognition tasks.

Table 1. X-vector DNN architecture [18]

Layer	Layer Context	Table Context	Input x Output
frame 1	[t-2, t+2]	5	120x512
frame 2	{t-2, t, t+2}	9	1536x512
frame 3	{t-3, t, t+3}	15	1536x512
frame 4	{t}	15	512x512
frame 5	{t}	15	512x1500
stats pooling	[0, T)	T	1500Tx3000
segment 6	{0}	T	3000x512
segment 7	{0}	T	512x512
softmax	{0}	T	512xN

3.2 SincNet Architecture

The SincNet architecture implements various band pass filters in its initial layers, that learns from raw audio signals. The SincNet filters are implemented using a set of mathematical operations which are stated below [14].
The function $h[.]$ is the Finite Impulse Response (FIR) filter used as a convolution filter.

$$y[n] \ = \ x[n] \ * \ h[n] \ = \ \sum_{l=0}^{L-1} \ x[l] * \ h[n \ - \ l] \tag{1}$$

$g[.]$ is a predefined function that depends on a few learnable parameters θ.

$$y[n] \ = \ x[n] \ * \ g[n, \theta] \tag{2}$$

$rect(.)$ is the rectangular function in the magnitude frequency domain and f_1 and f_2 are cut-off frequencies.

$$G[f, f_1, f_2] \ = \ rect\left(\frac{f}{2f_2}\right) \ - \ rect\left(\frac{f}{2f_1}\right) \tag{3}$$

Here, the *sinc* function is defined as $sinc(x) \ = \ \sin(x) \ / \ x$

$$g[n, f_1, f_2] \ = \ 2f_2 sinc(2\pi f_2 n) \ - \ 2f_1 sinc(2\pi f_1 n) \tag{4}$$

We have to ensure that $f_1 \geq 0$ and $f_2 \geq f_1$, therefore, the previous equation is actually fed with the following parameters:

$$f_1^{abs} \ = \ |f1| \tag{5}$$

$$f_2^{abs} \ = \ f1 \ + \ |f2 \ - \ f1| \tag{6}$$

Windowing is performed by multiplying the function with a window function w.

$$g_w[n, f_1, f_2] = \ g[n, f_1, f_2] * \ w[n] \tag{7}$$

The windowing function w is a hamming window which is given by $eqn(8)$.

$$w[n] = \ 0.54 \ - \ 0.46 * \ cos\left(\frac{2\pi n}{L}\right) \tag{8}$$

Once the filters are applied on the raw audio, we get features. These features can now be fed into any classifier. The vanilla SincNet architecture can be seen in Fig. 1. It takes raw audio signals as input, applies SincNet filters and then feeds it into a CNN model which is used as a classifier.

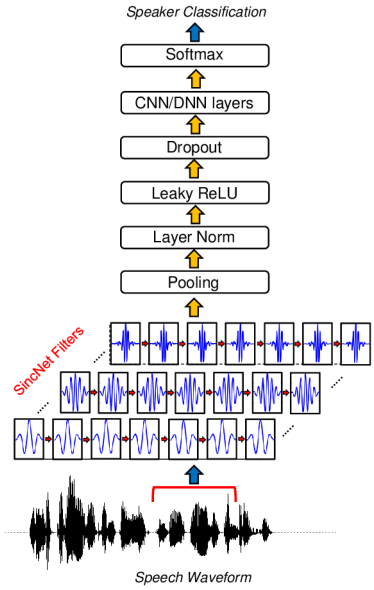


Fig. 1. This is the SincNet architecture. The image has been taken from the original paper [14].

3.3 Fusion Model

Our proposed fusion model fuses the pre-trained X-Vector with features extracted from the trained SincNet model. The concatenated features are fed to a fully connected dense layer. This is followed by two more fully connected dense layers as shown in Fig. 2. The idea behind using SincNet filters on the raw waveform is to extract more distinguishing features in the initial convolution

layers of the CNN architecture. X-Vectors are used to take advantage of the fact that this embedding is an efficient method to churn out fixed dimension features variable length speech utterances, something which is challenging in plain CNN techniques, making it efficient both in terms of speed and accuracy.

4 Experimental Setup

4.1 Dataset

We have carried out our experiments on the publicly available dataset VoxCeleb1 [11]. VoxCeleb is an audio-visual dataset consisting of short clips of human speech, extracted from interview videos uploaded to YouTube. We have used the raw audio files for our experiments. The VoxCeleb1 dataset consists of videos from 1,251 celebrity speakers. This means a size of 1,251 speakers and about 21k recordings.

Table 2. VoxCeleb1 dataset distribution [11].

	Dev	Test
Number of speakers	1,251	1,251
Number of videos	21,245	1,251
Number of utterances	145,265	8,251

4.2 Model Architecture

In order to carry out comparative results we have experimented with the original SincNet architecture and our proposed architecture. The original SincNet architecture uses SincNet filters for feature extraction whereas our architecture makes use of both SincNet and X-Vector. The classifier used in all the cases consists of several fully connected layers (DNN classifier) with softmax layer as the output layer. The models can be categorized as:

1. SincNet based feature extractor and DNN classifier.
2. X-Vector embedding and DNN classifier.
3. X-Vector and SincNet based feature extractor and DNN classifier *(Proposed)*.

Our proposed fusion model fuses the pre-trained X-Vector with features extracted from the trained SincNet model. The concatenated features are fed to a fully connected dense layers as shown in Fig. 2.The output obtained after convolution step is flattened and fed into the Dense Layer which is further concatenated with X-vector Embedding. The X-vector combined with the Dense layer constitutes fully connected layer 1 (FC1) which is further connected to FC2 and FC3. All the dense layers use Leaky ReLU as activation. The softmax layer is used as the output layer.

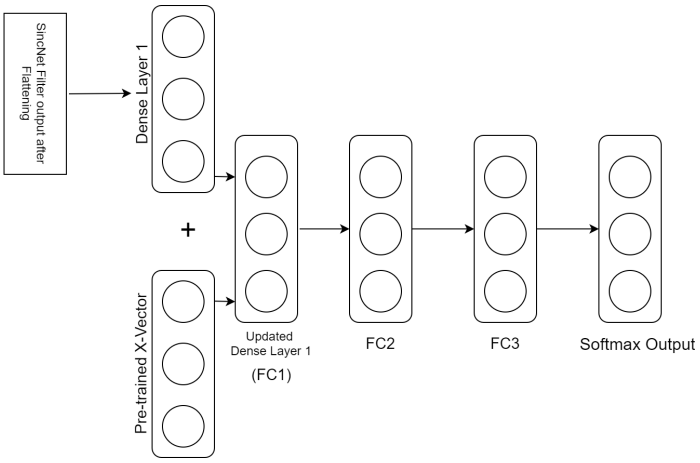


Fig. 2. The proposed fusion model.

5 Experimental Results

The experiments were carried out in Python 3 on a Nvidia *GeForce GTX 1080 Ti* GPU equipped with 16GB of memory. The software implementation of our code is made available online on GitHub [10]. The tests were carried out on the VoxCeleb1 dataset and the results are summarised in Table. 3. We calculated the Equal Error Rate (EER) for all the experiments and lower the value the of EER, the better is our model. The results were in alignment with our expectations which means that our system performed the best out of all the architectures tested.

We have an EER score 8.2 for pure SincNet based architecture, an EER score of 5.8 for the X-vector based architecture, and the best EER score of 3.56 using our SincNet and X-vector embedding based architecture. Our architecture proposed a framework which resulted in the EER score of 3.56 which was an improvement over the previous best EER score of 4.16 using X-vector over SITW core [18]. Fig. 3 shows a comparison of EER score of various architectures as the number of epochs increases. The proposed fusion model exhibits a consistently low EER over all epochs.

Table 3. Experimental Results

Architecture Used	Training Dataset	EER (On Test Data)
SincNet	VoxCeleb1	8.2
X-Vector	VoxCeleb1	5.8
Proposed	**VoxCeleb1**	**3.56**

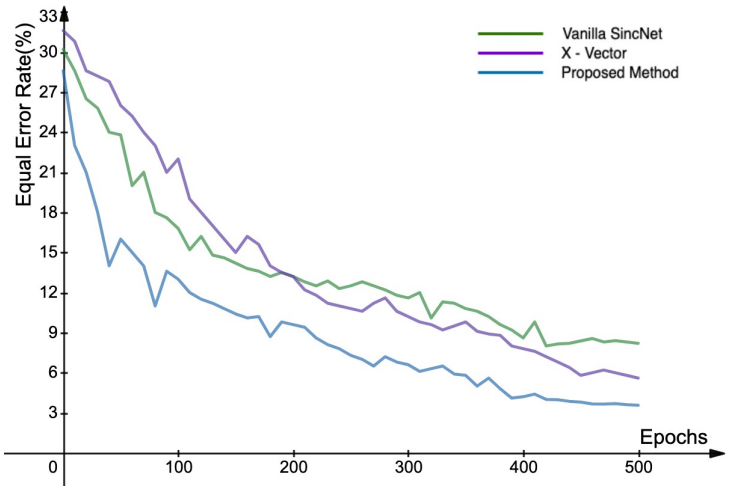


Fig. 3. Comparison of EER score of various architectures over epochs.

6 Conclusions

In this paper we propose a novel fusion model involving two successful deep architectures for speaker recognition:- SincNet and X-Vector.The features extracted from the two sources are fused by concatenation and learnt using fully connected dense layers. We achieved an increase in the EER score from current state-of-the-art by 14.5% on the VoxCeleb1 Dataset. It also showed quite a significant improvement over using vanilla SincNet which resulted in an EER score of 8.2. Further improvement over this architecture can be carried out by combining the VoxCeleb2 dataset along with VoxCeleb1 dataset and using noise removal techniques prior to feeding into the network.

References

1. Beigi, H.: Speaker recognition. Springer (2011)
2. Chung, J.S., Nagrani, A., Zisserman, A.: Voxceleb2: Deep speaker recognition. arXiv preprint arXiv:1806.05622 (2018)
3. Cumani, S., Plchot, O., Laface, P.: Probabilistic linear discriminant analysis of i-vector posterior distributions. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 7644–7648. IEEE (2013)
4. Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing **19**(4), 788–798 (2010)
5. Ghalehjegh, S.H., Rose, R.C.: Deep bottleneck features for i-vector based text-independent speaker verification. In: 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). pp. 555–560. IEEE (2015)

6. Huang, H., Sim, K.C.: An investigation of augmenting speaker representations to improve speaker normalisation for dnn-based speech recognition. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4610–4613. IEEE (2015)
7. Kenny, P.: Bayesian speaker verification with heavy-tailed priors. In: Odyssey. vol. 14 (2010)
8. Lee, J., Kim, T., Park, J., Nam, J.: Raw waveform-based audio classification using sample-level cnn architectures. arXiv preprint arXiv:1712.00866 (2017)
9. Li, C., Ma, X., Jiang, B., Li, X., Zhang, X., Liu, X., Cao, Y., Kannan, A., Zhu, Z.: Deep speaker: an end-to-end neural speaker embedding system. arXiv preprint arXiv:1705.02304 (2017)
10. Mayank Tripathi, D.: Icaisc-speaker-identification-system, `https://github.com/mayank408/ICAISC-Speaker-Identification-System`
11. Nagrani, A., Chung, J.S., Zisserman, A.: Voxceleb: a large-scale speaker identification dataset. arXiv preprint arXiv:1706.08612 (2017)
12. Novotnỳ, O., Plchot, O., Glembek, O., Burget, L., Matějka, P.: Discriminatively re-trained i-vector extractor for speaker recognition. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6031–6035. IEEE (2019)
13. Palaz, D., Collobert, R., et al.: Analysis of cnn-based speech recognition system using raw speech as input. Tech. rep., Idiap (2015)
14. Ravanelli, M., Bengio, Y.: Speaker recognition from raw waveform with sincnet. In: 2018 IEEE Spoken Language Technology Workshop (SLT). pp. 1021–1028. IEEE (2018)
15. Sainath, T.N., Weiss, R.J., Senior, A., Wilson, K.W., Vinyals, O.: Learning the speech front-end with raw waveform cldnns. In: Sixteenth Annual Conference of the International Speech Communication Association (2015)
16. Snyder, D., Garcia-Romero, D., Povey, D.: Voxceleb models, `http://kaldi-asr.org/models/m7`, accessed: 2019-11-30
17. Snyder, D., Garcia-Romero, D., Povey, D., Khudanpur, S.: Deep neural network embeddings for text-independent speaker verification. In: Interspeech. pp. 999–1003 (2017)
18. Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., Khudanpur, S.: X-vectors: Robust dnn embeddings for speaker recognition. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5329–5333. IEEE (2018)
19. Susan, S., Sharma, S.: A fuzzy nearest neighbor classifier for speaker identification. In: 2012 Fourth International Conference on Computational Intelligence and Communication Networks. pp. 842–845. IEEE (2012)
20. Variani, E., Lei, X., McDermott, E., Moreno, I.L., Gonzalez-Dominguez, J.: Deep neural networks for small footprint text-dependent speaker verification. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4052–4056. IEEE (2014)
21. Wyse, L.: Audio spectrogram representations for processing with convolutional neural networks. arXiv preprint arXiv:1706.09559 (2017)
22. Zhang, C., Koishida, K., Hansen, J.H.: Text-independent speaker verification based on triplet convolutional neural network embeddings. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP) **26**(9), 1633–1644 (2018)

STRENGTH AND DURABILITY PERFORMANCE OF FLY ASH BASED GEOPOLYMER CONCRETE USING NANO SILICA

Sameer Vyas [*1], Mohammad Sameer [2], Shilpa Pal [3], Neetu Singh [4]
[*1,4] Central Soil and Materials Research Station, New Delhi, India
[2] Department of Civil Engineering Gautam Buddha University, Greater Noida, U.P., India
[3] Department of Civil Engineering Delhi Technological University, Delhi, India

DOI: https://doi.org/10.29121/IJOEST.v4.i2.2020.73



ABSTRACT

With the increasing infrastructure development across the globe, the demand of cement production increases day by day. However, the production of cement is associated with the emission of large amount of CO2 causing global warming. Scientist and engineers are in search of a green ecofriendly alternative for concrete production. Geopolymers are rapidly emerging as an alternative to Portland cement as the binder of structural concrete. In this respect, the fly ash based geopolymers shows considerable prospect for application in concrete industry as an alternative binder to the Portland cement. Development of geopolymer concrete using class F fly ash brings many advantages like; enhancing workability, durability, better strength as well as lowering the price. There is not only a reduction in the greenhouse footprint but, also considerable increase in strength and resistivity to adverse conditions.

In order to enhance the performance of Geopolymer concrete, the use of Nano-silica is found to be suitable and practiced by researchers.

Use of Nano materials as fillers in the concrete matrix has proven effective in increasing mechanical and durability properties. This research is based on performance evaluation of geopolymer concrete using different percentage of Nano-silica. It was observed that Geopolymer concrete with Nano-silica (GPC-N) shows good compressive strength as well as durability under aggressive conditions.

The materials performance was also investigated using X-Ray Diffraction technique. (XRD). Results show that the presence of nanosilica enhanced the performance of Geopolymer concrete with respect to strength and durability purposes.

Keywords: Geopolymer; Fly Ash; Nano-Silica; Durability; Compressive Strength; Aggressive Conditions.

1. INTRODUCTION

Concrete made with Portland cement is the most widely used material on earth. The concrete industry is the largest user of natural resources in the world [1]. Significant increases in cement production have been observed and are anticipated to increase due to the massive increase in infrastructure and industrialization in India, China and South America. It is generally agreed that the production of Portland cement clinker is expensive and ecologically harmful. As one ton of OPC manufactured, nearly one ton of CO2 is produced depending on the production process adopted [2]. The emissions generated by Portland cement productions are principal contributors

to the greenhouse gas (GHG) effect. For instance, the production of Portland cement for concrete accounts for an estimated 5 percent of global anthropogenic carbon dioxide [3].

In view of the serious impact of carbon dioxide on the environment and the continued anticipated growth of industrialization and urbanization, there is a need to redirect the building industry away from its overwhelming reliance on Portland cement by developing alternative binder systems.

The two options which have attracted attention as alternative binders are (i) the partial replacement of cement by industrial byproducts like fly ash and slag and (ii) the use of geopolymer binders. The first alternative has been widely researched and abundant information on the fresh and hardened properties of concrete with partial replacement of cement has led to the use of such blended cements [4, 5, 6, 7, 8].

The second alternative, geopolymer binders, is an emerging area of technology. Davidovits [9] first proposed that an alkaline liquid could be used to react with the silicon (Si) and aluminum (Al) in a source material of geological origin or in by-product materials such as fly ash and rice husk ash to produce cementitious binders. Because the chemical reactions that takes place in this case is a polymerization process and the source materials are of geological origin, he coined the term„ geopolymer‟ to represent these binders. Geopolymers are members of the family of inorganic polymers. The polymerization process involves a substantially fast chemical reaction under alkaline condition on Si-Al minerals, that results in a three-dimensional polymeric chain and ring structure consisting of Si-O-Al linkage.[10]. Davidovits showed that higher amount of reactive silica resulted in the formation of higher amount of geo polymeric gel and consequently developed higher mechanical strength. He also reported that the low calcium fly ash based geopolymer concrete was emerging as an alternative low emission binding material compared to OPC concrete.

It has been found that geopolymer concrete has good engineering properties including compressive strength, enhanced tensile strength, bond strength, sulphate and acid resistance and the potential for enhanced durability [11, 12, 13].

Research based on geopolymer concrete has mostly been restricted to defining the strength and mix proportion variations [14-20]. Nano technology is having an important role in construction field. Supplementary cementitious materials have been a good choice for replacing cement and adding nanoparticles improves the performance of concrete. [21,22, 23]

Integration of nano silica with fly ash based geopolymer concrete has shown better results than the former. Nanotechnology being one of the most promising research fields in the growth and development of the modern industry has also been used to harness optimum results in the concrete industry by providing opportunities to improve the working of the concrete on molecular level and to enhance the properties of the materials. A dense formation of concrete could be achieved by stacking calcium silicate hydrate (5 nanometer diameter) molecules in pyramid to reach the highest density of spherical packing i.e. 74%. The density and hence the strength of the concrete can be substantially increased by imparting nano materials in the interstitial sites of the calcium silicate hydrate structure.

This paper is based on the research conducted for understanding the behavior and performance of geopolymer concrete using nano silica in different percentages. The mechanical strength analysis was done by compressive strength test using Compression Testing Machines (CTM). The durability properties based on chloride resistance test, sulphate resistance test, acid resistance test, residual alkalinity, fire resistance tests were also integrated. The mineralogical and micro structural assessment of nano silica modified geopolymer concrete properties was evaluated further by the use of X-RAY diffractometer (XDR).

2. MATERIALS AND METHODS

The materials used to determine the effect of Nano-silica on strength and durability of fly ash based Geo-polymer concrete specimens are low-calcium dry fly ash as the source material, aggregates, alkaline liquids, water, and nano-silica.

A. Fly Ash

Low calcium (Class F) dry fly ash obtained from N.T.P.C thermal power plant has been used as the base material. The chemical composition is presented in Table1.

B. Alkaline liquid

The alkaline liquid was prepared by mixing sodium silicate solution and sodium hydroxide solution (12 M). The sodium silicate solution contained 30% of solids and 70% of water.

C. Nano Silica

Nano silica of particle size 10 nm was obtained from Bee-chem Chemicals Ltd., Kanpur.

D. Aggregates

Locally available aggregates, comprising 10mm size were used and fine aggregate was standard sand which was well within the limit confirming zone--II as per IS Code 383.1970.

Table 1: Chemical composition of fly ash.

Oxides	Test method	Test result % by weight
Silica	IS:1727	60.22
Iron oxide	IS:1727	4.81
Alumina	IS:1727	26.26
$SiO_2+Al_2O_3+Fe_2O_3$	IS:3812(pt.1)	91.29
Clacium oxide	IS:1727	1.84
Magnesium oxide	IS:1727	1.71
Total sulphur	IS:1727	0.02
Available alkalies	IS:3812(pt1)	0.26
Loss on ignition	IS:1727	0.95

3. EXPERIMENTAL WORK

A. Mix Proportions

The mix proportions for this study was prepared as per standard procedures using nano silica in different proportions.

Density of concrete = 2400kg/m³

- (Flyash + Fine aggregates + Coarse aggregates + Alkaline solution + Nano silica).
- Ratio of Flyash : fine : coarse
- 1 : 1.13 : 2.25
- Taken, Alkaline to fly ash ratio = 0.45.
- Ratio of NaOH to Na_2SiO_3 = 2.5/
- NaOH: 12M.
- Nano silica added in different percentages. (0%, 2%, 4%, 6%, 8% by weight of flyash).

The final mix design for making Geopolymer concrete with nano silica (GPC-N) for $1m^3$ is given in Table 2.

Table 2: Final Mix Design

Materials	**Quantity (kg/ m³)**
Fly ash	496.55kg
Fine Aggregate	560 kg
Coarse Aggregate	1120 kg
NaOH (12 M)	51.71 Kg. solid 108 lit. of water
Sodium silicate Gel, containing 70% of water and 30% of solids.	63.84 Kg. (70% of water)
Nano silica	Nano silica added in different percentages i.e. 2%, 4%, 6%, 8% by weight of fly ash.

B. Final Mix Design Proportions for 1 cube

For the final mix design of GPC-N, the nano silica was added in predefined percentages (0%, 2%, 4%, 6%, 8%) by weight of fly ash. Total 6 Nos. of trial mix designs were casted for this research. 5 trial mix designs were rejected for dearth in strength and durability properties.

C. Casting of Test Specimens

Preparation of alkaline liquids

The alkaline liquid was prepared one day advance of the casting by mixing sodium hydroxide (NaOH) solution of 12 M with sodium silicate solution.

Casting of Fresh geo-polymer Concrete with nano-silica

The casting of cubes were done as per mix design approach by suitably mixing all ingredients in exact proportions using standard procedures.

The concrete slump before putting it into moulds was measured to be 92mm which is medium. The fresh concrete after being put into moulds was compacted properly.

The process of casting is presented in fig.1-2



Figure 1: Preparation of Mix.



Figure 2: Casting of Concrete cube.

D. Curing of Test Specimens

The test specimens were cured in the oven. The specimen were heated at 60° C for 72 hours to gain strength. The process of curing is presented in fig.3-4.





Figure 3, 4: Curing of Concrete Cubes.

After required curing the samples were left to air dry in the laboratory until the day of testing.

E. Strength Property

Compressive strength

The 70.6 x 70.6 mm geopolymer concrete (GPC-N)) cube specimens were prepared for each mix proportions of geo-polymer concrete with nano-silica. The specimens were tested for compressive

strength after 72 hours of casting.. The specimens made from final mix design were cured at 60°C for 72 hours, The compressive strength was measured on Compression Testing Machine (CTM).

The GPC-N Cubes are presented in fig.5 and Compression testing machine is presented in Fig. 6.



Figure 5: Cured GPC-N Cubes.



Figure 6: Compression testing machine.

F. Durability Test

Chloride Resistance

Marine structures are highly subjected to chloride attack and due to the chloride penetration, the reinforcement is subjected to corrosion. The test specimens of GPC-N were immersed in 10% sodium chloride solution for the period of 4-8 week exposure. The performance of GPC-N cubes were evaluate on the basis of Change in Compressive strength.

Sulphate Resistance

Sulphate usually present in ground water or soil which comes into the contact of concrete and affects it. The test specimens were immersed in 10% Sodium sulphate (Na_2SO_4) solution. The performance of GPC-N cubes after exposures of 4-8 weeks were evaluate on the basis of Change in Compressive strength.

The exposure of GPC-N cubes in Chloride & Sulphate ion solution is presented in fig. 7.



Figure 7: Exposure of GPC-N cubes in Chloride & Sulphate ion solution.

Acid Resistance

The acid resistance test was performed to study the effect of Sulphuric acid (H_2SO_4) and Hydrochloric acid (HCl) on GPC-N specimens. The test specimens of GPC-N were immersed in 3% & 5% solutions of concentrations of sulfuric and hydrochloric acid for 4-8 weeks exposure. The performance of GPC-N specimens were than evaluate on the basis of Change in compressive strength.

G. X-Ray Diffraction Test

X-ray diffractometer test is a non-destructive testing used for the characterization and identification of crystalline materials.

The samples for XRD test were prepared by crushing the GPC-N concrete cube particles and passed through 100-micron sieve. The samples of flyash, Nano-silica, and crushed cubes samples after test of compressive strength and durability were taken to study XRD pattern.

4. RESULTS AND DISCUSSIONS

4.1. Compressive Strength

The results for compressive strength for Final mixed design is presented in table 3 and Fig.8. With the increase in the percentage of nano silica there is an appreciative increase in the strength up to 8%. Addition of nano silica beyond 8% decreases the strength by makes the concrete brittle.

Table 3: Compressive Strength for Final Mix Design with Different Percentage of Nano silica.

	Compressive strength in (MPa)					
	Nano silica 0%	**Nano silica 2%**	**Nano silica 4%**	**Nano silica 6%**	**Nano silica 8%**	**Nano silica 10%**
Final mix design	24.0	26.7	29.8	32.0	35	31.7

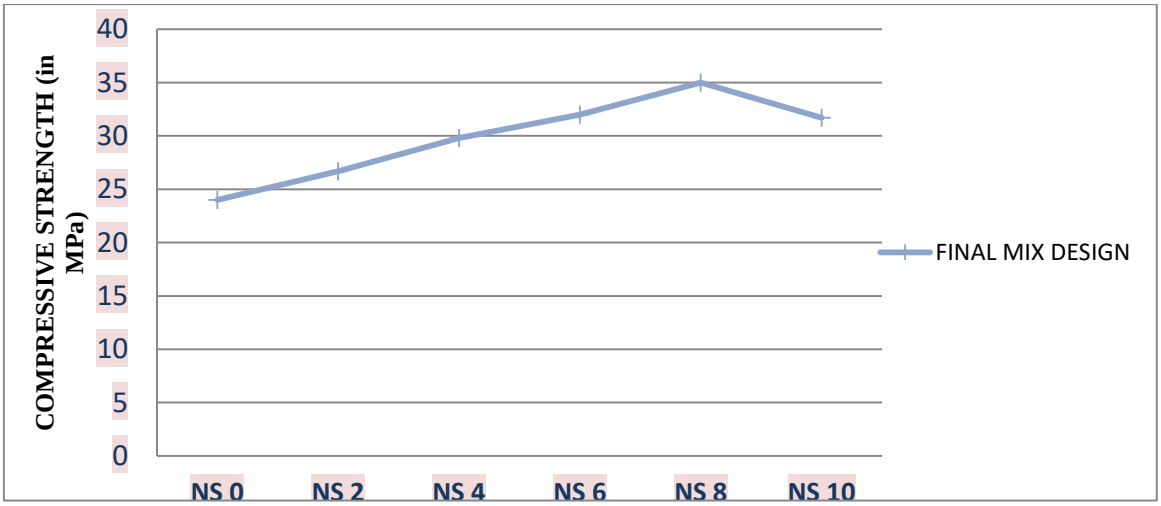


Figure 8: Variation in Compressive Strength with Nano silica percentages.

4.2. Durability Tests

Effect of Chloride Ion Exposure on Compressive Strength of GPC-N Cubes.
The compressive strength of these test specimens without any exposure of chloride solution was taken as the initial (reference) compressive strength. The test results for 4-8 weeks exposure periods are presented in Fig.9.

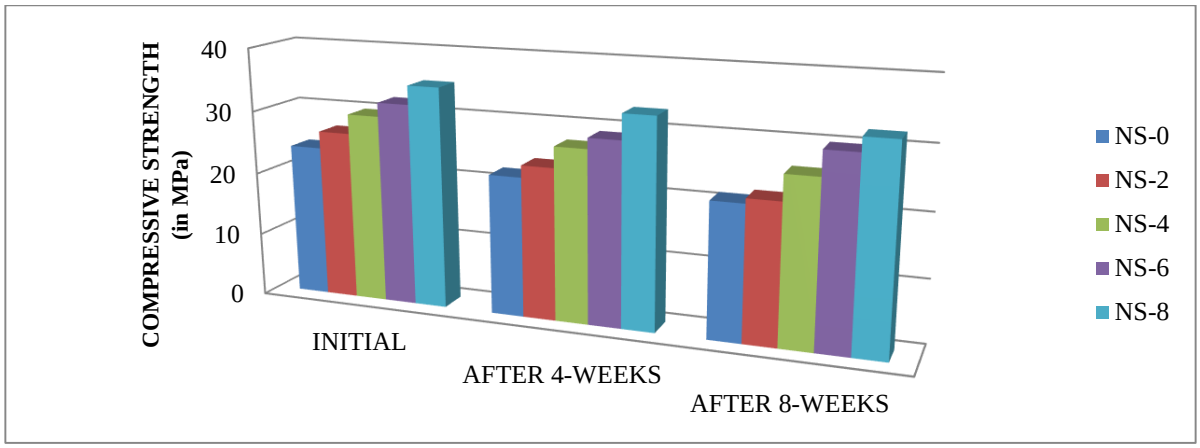


Figure 9: Compressive Strength after Exposure of GPC-N cubes in Chloride Solution.

These test results for chloride resistance depict that the GPC-N cubes exposed to 10% sodium chloride solution for 4-8 weeks shows very little change in compression strength. Therefore, geopolymer concrete is shows remarkable resistance to chloride exposure.

Effect of Sulphate Ion Exposure on Compressive Strength of GPC-N Cubes.

The test results for initial reference and 4-8 weeks exposure periods are presented in Fig.10

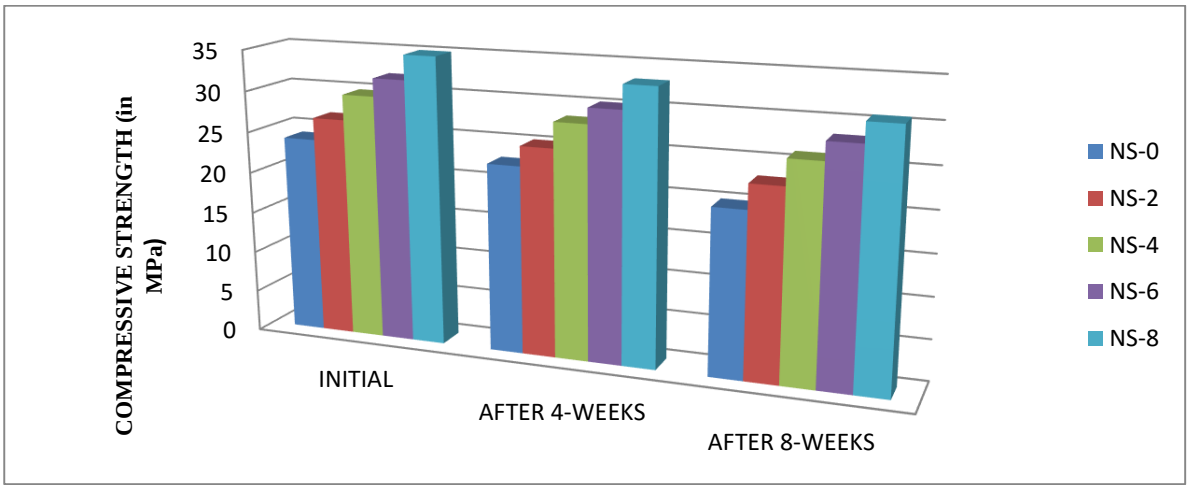


Figure 10: Compressive Strength after Exposure of GPC-N cubes in Sulphate Solution.

Very small change was observed after the exposure of GPC-N cubes in 10% sodium sulphate solution. However, GPC-N with 8% Nano silica shows good resistance to Sulphate exposure.

Change in Compressive Strength After Exposure in Acid Solutions

The change in compressive strength of GPC-N test specimens exposed to 5% solution of sulphuric and hydrochloric acid solution are presented in fig.11 & 12.

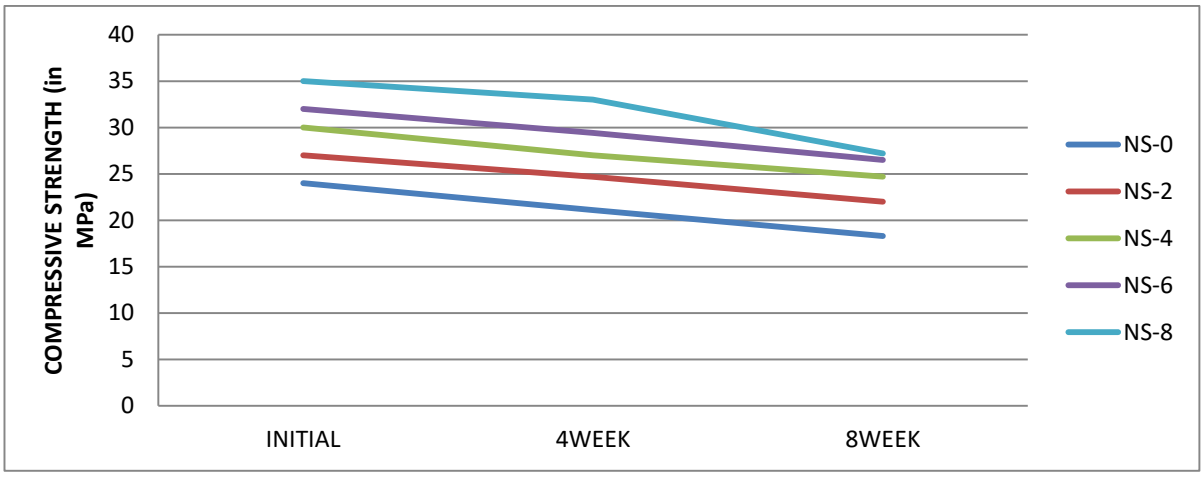


Figure 11: Change in Compressive Strength of GPC-N

Specimens Exposed to 5% Sulfuric Acid Solution.

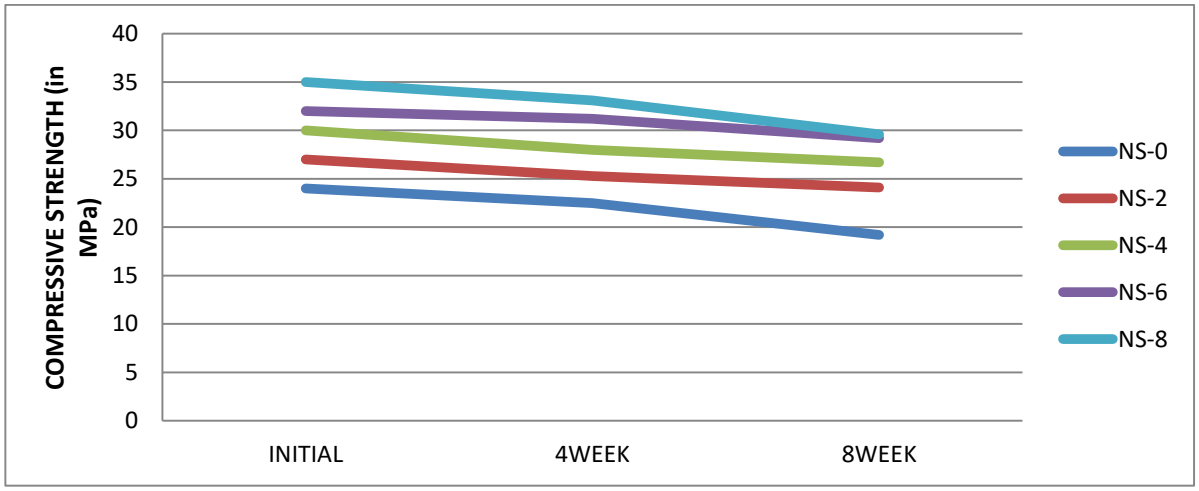


Figure 12: Change in Compressive Strength of GPC-N Specimens Exposed to 5% Hhydrochloric Acid Solution.

The plain geo-polymer concrete specimen shows 20% reduction in strength whereas in GPC-N, 16% reduction in compressive strength was observed which suggested that the GPC-N specimen having good acid resistance. The concentration of acid and period of exposure are the key parameters which are responsible for degradation in strength property.

4.3. Mineralogical Characteristics

The Fig:13 represents XRD spectra obtained for Fly ash and Plain GPC.

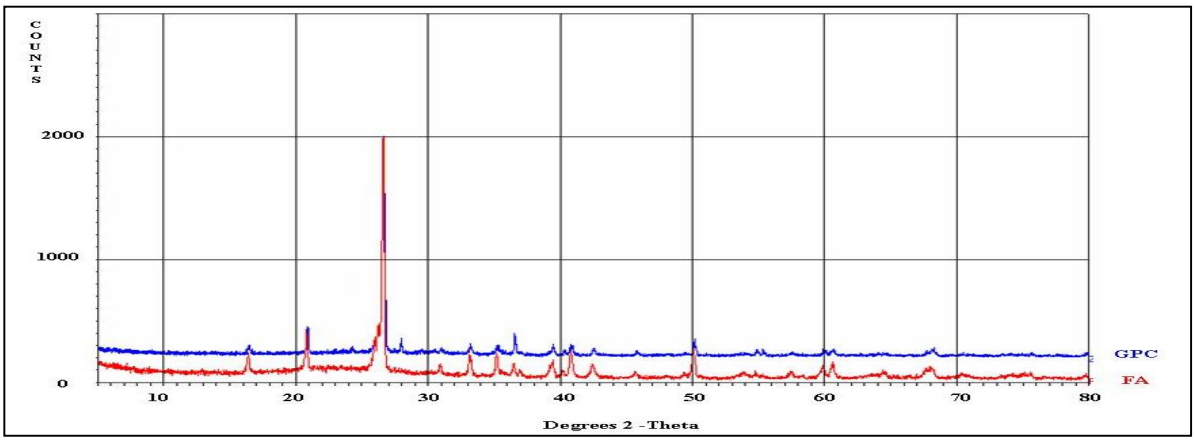


Figure 13: XRD pattern of fly ash and GPC

The pattern of XRD analysis shows that besides crystalline phase composed of quartz and mullite, Flyash is primarily composed of amorphous material. On geo-polymerization process the peaks represents the amorphous phase of fly ash appears slightly shifted towards higher theta values which shows the transformation of fly ash amorphous phase into new geo-polymer concrete amorphous phase. These new diffraction peaks are mainly due to the formation of new crystalline zeolitic phases such as hydroxysodalite, analchime and herschelite. Fig 14 represents the XRD pattern of GPC-N specimens.

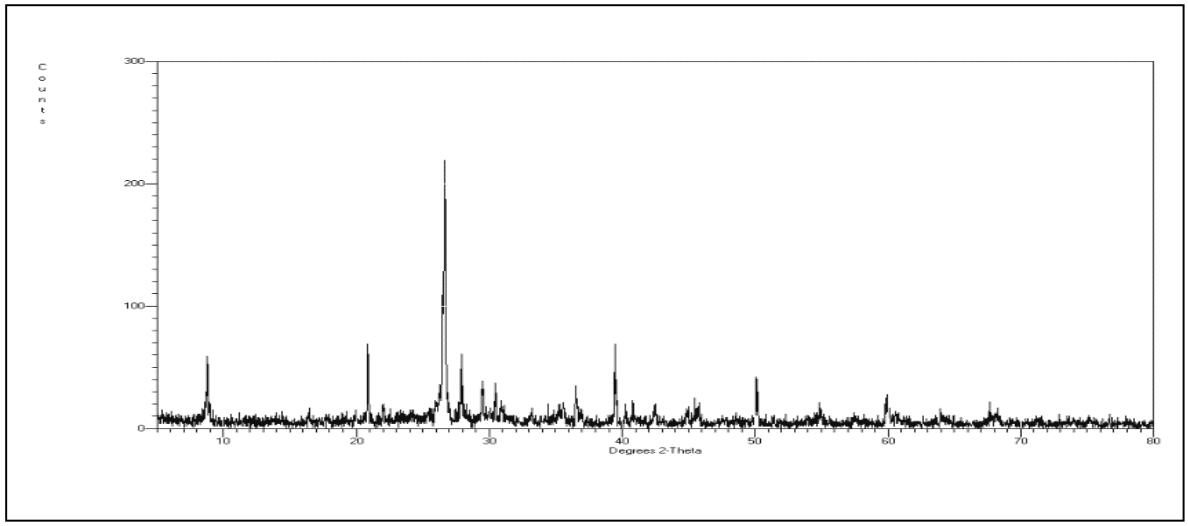


Figure 14: XRD pattern of GPC-N

The XRD analysis pattern shows that the geopolymer concrete with 8% nano silica addition, some specific extra peak positions are observed. It indicates the formation of new crystalline phase as compared to the geo-polymer concrete without nano-silica samples.

5. Conclusions and Recommendations

The following conclusions have been drawn from the study of effect of Nano-silica on the strength and durability of fly ash-based geopolymer concrete. The GPC-N cubes were casted with five

different percentages of nano silica i.e. 0%, 2%, 4%, 6%, and 8% by weight of fly ash respectively. The outcome of the different tests conducted on these GPC-N concrete cubes are summarized as follows: -

1) The compressive strength of GPC-N increased with increase in the nano silica percentage up to 8% due to low permeability and dense geopolymer concrete matrix, as compared to geopolymer concrete without nano silica, at controlled curing temperature and curing period.
2) The chloride and sulphate intrusions were minimum in GPC-N cubes with 8% nano silica. The rate of decrease of compressive strength in GPC-N decreased with the addition of nano silica. Tests also indicated that there was no formation of ettringite or gypsum from the main products of polymerization in GPC-N.
3) The GPC-N cubes with 8% nano silica showed fair resistance to acid environment as the decrease in compressive strength was marginal. With the increase in HCl and H_2SO_4 concentration there was a decrease in the compressive strength and rate of surface erosion, however the acid resistance was better than that of geopolymer concrete without nano silica.
4) The investigations so far have indicated that geopolymer concrete with nano silica addition have found potential replacement for the OPC in concrete. The GPC-N also performed well in different environments.

Acknowledgements

The authors are extremely thankful to all the officers and staff of concrete technology division for extending laboratory facility and arrangement of materials required for testing work.

References

[1] Mehta, P. K. “Greening of the Concrete Industry for Sustainable Development”, ACI Concrete International, 24(7), 2002, pp.23-28.
[2] McLellan BC, Williams RP, Lay J, et al. Costs and carbon emissions for geopolymer pastes in comparison to ordinary Portland cement. J Clean Prod 2011, 19: 1080–1090.
[3] Karstensen, K.H., “Cement Production in Vertical Shaft Kilns in China –Status and Opportunities for Investment,” UNIDO Contract RB-308-D40-8213110-2005, 2006.
[4] Mehta, P. K. “High Performance, High Volume Fly Ash Concrete for Sustainable Development” in Proceedings of International Workshop on Sustainable Development and Concrete Technology, Beijing, China, 2004, pp. 3-14.
[5] Stevenson, M. and Panian, L. “Sustainability through Strength” Concrete International, 31(3), 2009, pp. 34:39.
[6] Poon, C. S., Lam, L. and Wong, Y. L. “A Study on High Strength Concrete Prepared with Large Volumes of Low-Calcium Fly Ash”, Cement and Concrete Research, 30(3), 2000, pp.447-455.
[7] Papadikis, V. G. “Effect of Fly Ash on Portland Cement Systems, Part I: Low-Calcium Fly Ash” Cement and Concrete Research, 29(11), 1999, pp. 1727-1736.
[8] Malhotra, V. M. “High-Performance High-Volume Fly Concrete”, Concrete International 24(7): 2002, pp. 30-34.
[9] Davidovits, J. “Geopolymers: inorganic geopolymeric new materials” Journal of Thermal Analysis, 37, 1991, pp. 1633-1656.
[10] Davidovits J. Chemistry of geopolymeric systems, terminology. In Proceedings of the 2nd International Conference on Geopolymer, Saint-Quentin, France, 1999: 9–39.

[11] Sofi, M., van Deventer, J. S. J., Mendis, P. A. and Lukey, G. C. "Bond performance of Reinforcing Bars in Inorganic Polymer Concrete (IPC)", Journal of Materials Science, 42(9), 2007, pp. 3107-3116. 65.
[12] Sofi, M, van Deventer J, S. J., Mendis, P, A, and Lukey G. C. "Engineering Properties of Inorganic Polymer Concretes (IPCs)" Cement and Concrete Research, 37(2), 2007, 251-257.
[13] Provis, J. L., Muntingh, Y., Lloyd, R. R., Xu, H., Keyte, L. M., Krivenko, P. V. and van Deventer, J. S. J "Will Geopolymers Stand the Test of Time?" Ceramic Engineering and Science Proceedings, 28(9), 2007, 235-248.
[14] Pal S, Mandal S. Different thermal activation effect on fly ash based geopolymer concrete. Indian ConcrInst J 2011; 12:23–5.
[15] Daniel LYK, Jay GS. Effect of elevated temperatures on geopolymer paste, mortar and concrete. CemConcr Res 2010; 40:02.
[16] Chindaprasirta P, Rattanasakb U, Taebuanhuadb S. Role of microwave radiation in curing the fly ash geopolymer. Adv Powder Technol 2013; 24:03.
[17] Peigang H, Dechang J, Meirong W, et al. Thermal evolution and crystallization kinetics of potassium-based geopolymer. Ceram Int 2011; 37:01.
[18] Hussin MW, Bhutta MAR, Azreen M, et al. Performance of blended ash geopolymer concrete at elevated temperatures. Mater Struct 2014.*http://dx.doi.org/10.1617/s11527*- 014-0251-5.
[19] Vijai K, Kumutha R, Vishnuram BG. Effect of types of curing on strength of geopolymer concrete. Int J PhysSci 2010;5(9).
[20] Guo X, Shi H, Dick WA. Compressive strength and microstructural characteristics of class C fly ash geopolymer. Cement Concr Compos 2010; 32:2.
[21] T.R.P raveenkumar, M.M. Vijayalakshmi, Effect of Nano particles on the properties of concrete, International journal of chemtech, Vol.8, No.7, 2015, pp 50-55.
[22] T.R. Praveenkumar , M.M. Vijayalakshmi , M.S. Meddah, Strengths and durability performances of blended cement concrete with TiO_2 nanoparticles and rice husk ash, Construction and Building Materials 217 (2019) 343–351.
[23] KotiChiranjeevi, M.M. Vijayalakshmi, Praveenkumar T R Behaviour of Fly Ash based Geopolymer Concrete using Nano-Material International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-12, October 2019,4436-4439.

*Corresponding author.
E-mail address: samyog78@ yahoo.com

Volume 8 Issue 1 (2020) 26-29 ISSN 2347 - 3258
International Journal of Advance Research and Innovation

Study on Adequacy of Functional Characteristics of a Typical Urban Waste Water Treatment Plant

Tarun Shokeen, SK Singh, Geeta Singh
Department of Environmental Engineering, Delhi Technological University, Delhi-42, India.

Article Info
Article history:
Received 15 February 2020
Received in revised form
29 February 2020
Accepted 8 March 2020
Available online 15 March 2020

Keywords Activated Sludge Process, Sewage Treatment Plant, MLSS, Adequacy

Abstract: This study examines the presence of pollution of river Yamuna in the city of Delhi. The condition of river Yamuna is deteriorating day by day. The main cause of this is the poor standard treated sewage from the sewage treatment plants being dumped into drains which directly meet into river Yamuna. In this study I audited Keshopur 40 MGD sewage treatment plant which works on activated sludge process in which Mixed Liquor Suspended Solids (MLSS) is the prime concern for the proper functioning of plant and checked the adequacy of the various functional units by comparing the detention time of each and every unit with the standard detention time given by code. Activated Sludge Process involves aeration and a biological flock composed of bacteria and protozoa to convert non settle particles into the settle one.ASP provides one of the highest degree of treatment within the limited cost involved. In this study various recommendations are also given for increasing the efficiency of sewage treatment plant.

1. Introduction

Delhi collects its sewage from a wide network of 7000 km of sewage lines across the city from the approved colonies, re-settlement colonies, urban villages, unauthorized-regularized colonies and unauthorized colonies. Laying of sewerage system in the unsewered areas is a crucial step for pollution abatement from river Yamuna. At present, Delhi Jal Board is supplying around 830 MGD of drinking water to population of Delhi. The sewage generated @ 80% is as per Central Public Health and Environmental Engineering Organisation (CPHEEO) norms and is estimated about 680 MGD.

There are 36 Sewage Treatment plants at 21 locations in Delhi. Municipal wastewater treatment in Delhi is undertaken through activated sludge process which is the most common suspended growth process. Activated sludge plant involves wastewater aeration in the presence of a microbial suspension, solid-liquid separation following aeration, discharge of clarified effluent, wasting of excess biomass, and return of remaining biomass to the aeration tank.

In activated sludge process wastewater containing organic matter is aerated in an aeration basin in which micro-organisms metabolize the suspended and soluble organic matter. Part of organic matter is synthesized into new cells and part is oxidized to CO_2 and water to derive energy. In activated sludge systems the new cells formed in the reaction are removed from the liquid stream in the form of a flocculent sludge in settling tanks. A part of this settled biomass, described as activated sludge is returned to the aeration tank and the remaining forms waste or excess sludge.

1.1 2.Objective of the Study

- To measure the all the functional units of keshopur 40 MGD sewage treatment plant in Delhi and compare them to the standards
- To give justifying recommendations for its proper functioning.

2.Literature review

Dubey and Gupta, Audit Process on Water Treatment Plant [1]. In this paper they discussed the energy audit process of the 42 MLD sewage treatment plant situated in Jabalpur area at lalpur (M.P). They found that more energy is consumed in Electric motors, pumps, lighting, wastewater etc. A large amount of electrical energy is consumed in induction motor used in Industry. In this paper they describe the auditing process and various factors affecting the efficiency of any equipment and how to calculate the exact values from the electricity bills, we are discussing use of energy per day - savings measures which energy auditors frequently use.

***Corresponding Author,**
E-mail address: tarunshokeen@yahoo.in; sksinghdce@gmail.com; geeta.singh@dce.ac.in

All rights reserved: http://www.ijari.org

The auditing process includes preliminary audit and detailed energy audit as-

1. Conducted meeting
2. Collected plant data
3. Conducted field investigation
4. Create Equipment Inventory
5. Follow up

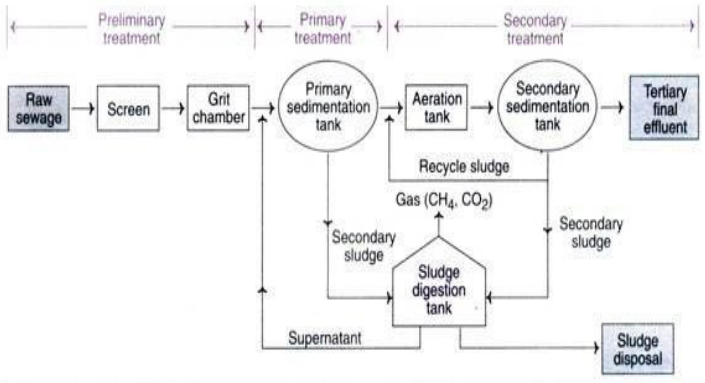


Fig. 1: Schematic showing Activated Sludge Process Sewage Treatment Plant.

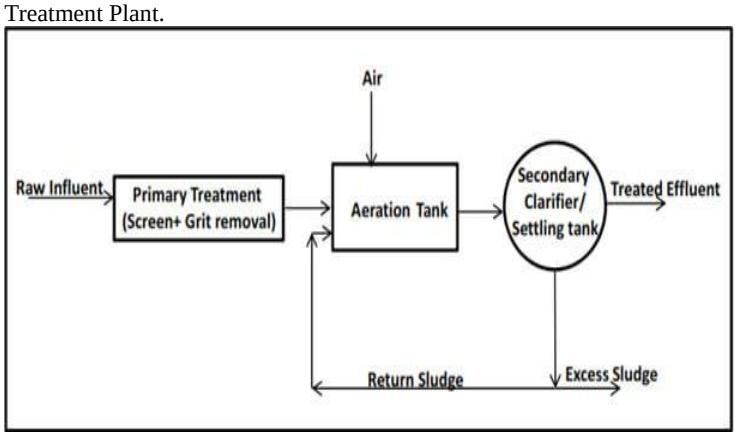


Fig.2: Schematic showing Extended Aeration Type Sewage Treatment Plant.

Hamed and Sabzali Comparative study of SMBR and extended aeration activated sludge processes in the treatment of high-strength wastewaters [2]. This study show the comparison between performance of extended aeration activated sludge (EAAS) and submerged membrane bioreactor (SMBR) systems in the treatment of wastewater under the same condition. The chemical oxygen demand (COD) concentration of the influent wastewater for the EAAS and SMBR systems was adjusted between 500–2700 and 500–5000 mg/L, respectively. Results showed that the quality effluent produced in SMBR system is much better than EAAS system in terms of COD, biochemical oxygen demand (BOD5), total suspended solids (TSS) and ammonium. By increasing the concentration of COD, the

concentration of mixed liquor suspended solids (MLSS) and the removal efficiency of organic matter in the SMBR system were increased regularly; however, the removal efficiency of COD in the EAAS system was irregular.

3. Materials and Methedology

3.1 Study Area

This study includes the functional inspection of the Kesopur 40MGD sewage treatment plant which is situated in New Delhi and measuring the dimensions of each unit and then calculating the detention time of each unit by dividing the volume of each tank by the flow rate present at the sewage treatment plant. Then this calculated detention time is compared with standard detention time given by codes.

Date & Time	: -	20.02.2019 & 2:15pm
Capacity	: -	40 MGD
Flow rate at the plant	:-	52.9 MGD
% Utilization	:-	132.25
Latitude	: -	28.6518
Longitude	: -	77.0809
Process	: -	ASP
Guiding Person	: -	KP Tiwari
Operator	: -	DJB

3.2 Screens

- Purpose- Removal of floating material.
- Total no. of screens –3(mechanical screens)
- None of the screen was working.

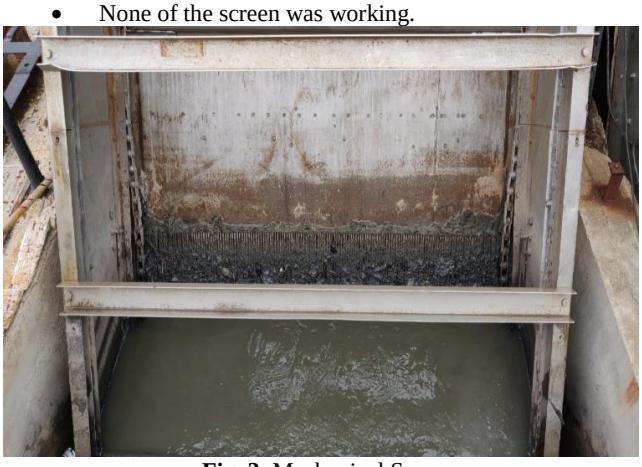


Fig. 3: Mechanical Screens

3.3 Degritting Unit

- No. of grit chamber = 3
- Length = 12.8m
- Breadth = 13.1m
- Height = 0.95m
- Flow = 10006.916m^3/hr = 166.782m^3/min.
- Flow at one unit = 55.59m^3/min.
- Detention time =2.86 min.
- Disposal of grit was proper.
- Grit chamber working satisfactorily.

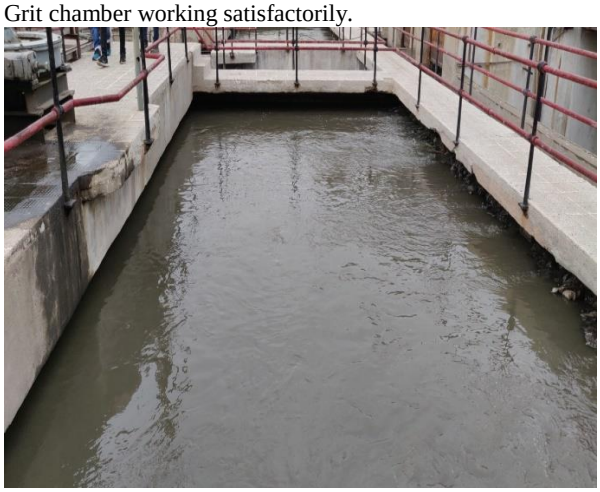


Fig. 4: Grit chamber

3.4 Primary Clarifier/PST

- No. of unit = 4
- Radius = 19.4m.
- Depth at centre = 4.8m.
- Depth at corner = 3.6m
- Average depth = 4.2m
- Volume = 4965.953m^3
- Flow = 10006.916m^3/hr
- Flow at one unit = 2501.729 m^3/hr.
- Detention time = 1.98hr.
- Some floating particles were observed in PST.



Fig. 5: PST

3.5 Aeration Unit

- No. of unit – 24 (13 working) +12 diffusers
- Volume – 77.4*93*6m^3
- Flow =10006.916m^3/hr.
- Detention time =4.31hr
- Sample was taken frequently to determine MLSS concentration.
- Diffused aeration system + mechanical aerators
- Colour of return sludge not dark brown.
- MLSS was very low = 740mg/l.(MLSS should be between 1500-3000mg/l)



Fig. 6: Aeration Tank

3.6 Secondary Sedimentation Tank

- No. of unit :- 4
- Radius :- 23m
- Depth at centre :- 5.3m
- Depth at corner :- 3.5m
- Average depth :- 4.4m
- Volume :-7312.37 m^3
- Flow :- 10006.916m^3/hr
- Flow at one unit :- 2501.729 m^3/hr.
- Detention time :- 2.92hr.



Fig. 7: Final Sedimentation Tank

3.7 Sludge Drying Beds

- No. of beds :- 46
- Floating material observed in drying beds.

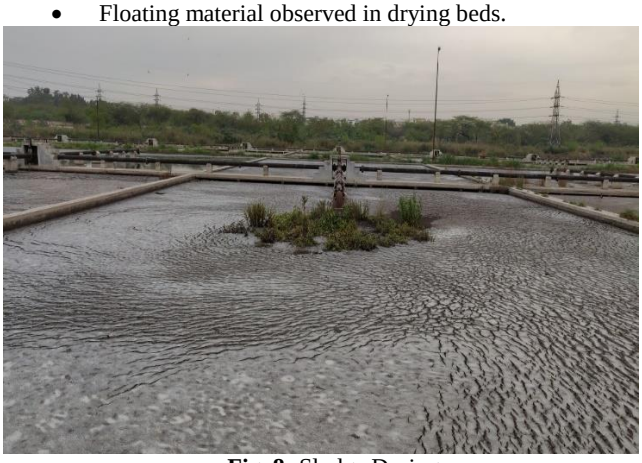


Fig. 8: Sludge Drying

3.8 Sludge Digestors

- No. of unit = 10.



3.9 Gas Holders

- No. of unit = 3
- Total gas compressors = 20
- All Gas formed was flared.

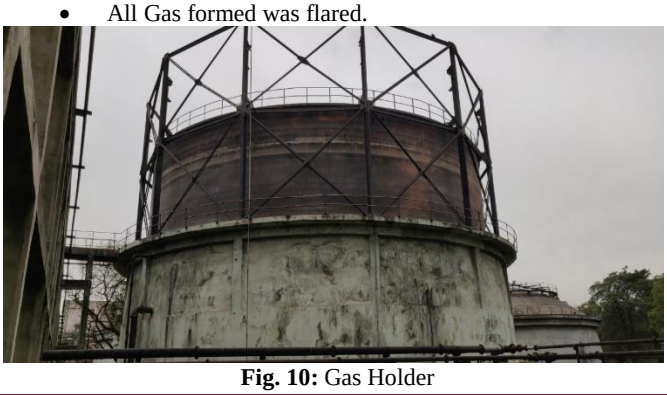


Fig. 10: Gas Holder

4. Results and analysis

Type of Unit	Measured/Calculated Values	Theoretical Recommended Values	Remarks
Screens	Bar spacing – 20mm Mechanical screens – 3 Manual screens – 0	Spacing : 20mm	All screens were not working .
Grit Chamber	Length-12.8m Breadth-13.1m Height-0.95m Discharge-166.782m^3/min. Flow at one unit - 55.59 m^3/min. Detention time-2.86min.	Detention time-40-60 seconds	Detention time is too high.
Primary Sedimentation Tank	Diameter- 39m Depth- 4.2m Discharge-10006.916m3/hr. Flowat one unit - 2501.729 m3/hr Detention time-1.98hrs	Detention time-1.5hr – 3hr	Detention time is withing the limits.
Aeration Tank	Length- 77.4m Breadth- 93m Depth- 6m Detention time-4.31hr MLSS -740mg/l	Detention time- 4-8hrs MLSS – 1500mg/l to 3000mg/l	Detention time is upto the mark and Sample was taken frequently for measuring MLSS concentration
Final Sedimentation Tank	Diameter- 46m Depth- 4.4m Discharge-10006.916m3/hr. Flowat one unit - 2501.729 m3/hr Detention time-2.92hrs	Depth-3.5-4.5m Detention time- 2-4hrs	Detention time is within the range.

5. Conclusions

The conclusions drawn from the present study were

- Screens should be made operational as this is the first unit of the STP responsible for the removal of floating particles in the sewage.
- Discharge should be increased to reduce the detention time in the grit chamber.
- Detention time of the primary sedimentation tank was within the limit.
- Detention time of aeration tank was also within the limits but the MLSS was too low for the proper functioning of the plants. So, the percentage of sludge from the FST should be increased to the aeration tank.
- Detention time of the final sedimentation tank was also within the range.

Thus, Screens should be made operational as floating particles were observed in almost all units. MLSS is too low therefore the quantity of sludge from FST that is to be transferred to inlet or at the aeration tank should be increased. Aesthetic view of plant should be enhanced. Gas generated should be used in electricity generation. Complete Rehabilitation is required which includes high design parameters, sludge management and power generation.

References

[1]. http://www.wefnet.org/MOPNEW/Operation_of_Municipal_Wastewater_Treatment_Plants/Chapter 22 Revised_6th Edition.pdf
[2]. A Barwal, R Chaudhary. To study the performance of biocarriers in moving bed biofilm reactor (MBBR) technology and kinetics

of biofilm for retrofitting the existing aerobic treatment systems: a review. Reviews in Environmental Science and Bio/Technology.
[3]. B Dubey, B Gupta. Audit Process on Water Treatment Plant.
[4]. SJ Burrian. The Historical Development of Wet-Weather Flow Management. US Environmental Protection Agency (EPA). National Risk Management Research Laboratory, Cincinnati, OH. Document No. EPA/600/JA-99/275, 1999.
[5]. GA Burton, RE Pitt, Robert. Chapter 2. Receiving Water Uses, Impairments, and Sources of Stormwater Pollutants. Stormwater Effects Handbook: A Toolbox for Watershed Managers, Scientists, and Engineers. New York: CRC/Lewis Publishers. ISBN 978-0-87371-924-7, 2001.
[6]. Chapter 3. Biological Treatment Processes. Emerging Technologies for Wastewater Treatment and In-Plant Wet Weather Management (Report). EPA. March 2013. EPA 832-R-12-011.
[7]. Chapter 3. Flow Equalization. Process Design Manual for Upgrading Existing Wastewater Treatment Plants (Report). EPA. October 1971.
[8]. EPA. Washington, DC . Primer for Municipal Waste water Treatment Systems. Document no. EPA 832-R-04-001, 2004.
[9]. Huber Company, Berching, Germany. Sedimentation Tanks. Archived 2012-01-18 at the Wayback Machine , 2012
[10]. SM Khopkar. Environmental Pollution Monitoring And Control. New Delhi: New Age International. p. 299. ISBN 978-81-224-1507-0, 2004.
[11]. Sanitation Systems – Sanitation Technologies – Activated sludge. SSWM. 27 April 2018.
[12]. G Tchobanoglous, FL Burton, HD Stensel. Metcalf & Eddy, Inc. Wastewater Engineering: Treatment and Reuse (4th ed.). McGraw-Hill. ISBN 978-0-07-112250-4, 2003.

Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020)
IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2

Synergism of Recurrent Neural Network and Fuzzy Logic for Short Term Energy Load Forecasting

Srishti Sharma[1], Rahul Kumar Agrawal[2],
[1]Dept. of Electrical Engineering,
[2]Dept. of Electrical Engineering,
Delhi Technological University, New Delhi, India
[1]srishtish21@gmail.com, [2]rahulagrawal0048@gmail.com

Madan Mohan Tripathi
Dept. of Electrical Engineering, Delhi Technological University, New Delhi, India
mmtripathi@dce.ac.in

2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC) 978-1-7281-4889-2/20/$31.00 ©2020 IEEE 10.1109/ICCMC48092.2020.ICCMC-00033

Abstract— **Utility companies need to plan the generation capacity and resources for establishing an efficient energy management with reduced wastage. Load forecasting would become essential to a power system's scheduling operations. Despite comprehensive work on short-term forecasting of loads using multiple machine learning models, there is room for improvement in prediction accuracy. A novel method of short-term load forecasting based on the combination of fuzzy logic with recurrent neural network (RNN) model has been proposed in this paper. The Fuzzy logic model maps the highly nonlinear relationship between weather parameters like changes in temperature and their effect on the regular demand for electric charges. The Fuzzy logic model maps the highly nonlinear relationship between weather parameters such as temperature changes and their effect on the daily demand for electric loads. The proposed approach combines the advantages of fuzzy logic and neural networks to predict the next day's load. Dataset of demand for electricity charging for a period of two years from 2013 to 2014 was obtained from ISO New England with a one-hour resolution. It is observed that in terms of precision, the Fuzzy-RNN hybrid model outperforms its counterpart RNN. It is observed that in terms of precision, the Fuzzy- hybrid model outperforms its counterpart RNN. The proposed model was contrasted with other state-of - the-art methods for short-term load forecasting including artificial neural network (ANN), fuzzy-ANN, support vector machine (SVM), and generic neural regression network (GRNN). Overall, the computed results conclude that the synergetic use of fuzzy logic with RNN model is successful in achieving higher accuracy by efficiently mapping the effect of weather parameters with a change in load demand. Fuzzy-RNN has performed best with the highest accuracy in load forecasting amongst the six models considered.**

Keywords— ***Load Forecasting, Fuzzy Logic, Generalized Regression Neural network***

I. INTRODUCTION

The widespread electricity supply industry has undergone structural and systematic changes in recent years. There are certain characteristics of electricity which makes it distinct from other commodities; storage of huge amount of electricity being economically inefficient, maintenance of demand-supply balance and transmission congestion problem. The conventional grid could not provide maximum return on investments to the power utilities and help consumers save on electricity consumption simultaneously. To ensure proper operation of power system, accurate estimation of electric load is required. Load forecasting is a way of estimating the future demand of load based on the previous data of the system. Load power forecasting has become necessary for planning the scheduling process of a power system at the generation end [1]. It greatly influences the financial performance of participants in electricity markets. Load forecasting is of three types: short term load forecasting (up to a few days), medium term load forecasting (from few weeks up to a year) and long term load forecasting (over a year and more). Short term load forecasting is performed for power system processes like load dispatching, unit commitment and power distribution whereas long term load forecasting have applications like setting up new power systems, purchasing energy and maintenance of the system. Short term load forecasting requires highly accurate instant results in a real time environment.

Attributed to the non-linear association between the parameters on which the electricity load depends, traditional statistical load forecasting methods prove to be incompetent. Application of artificial intelligence models for load forecasting has been in research since the late 1980's. Among various artificial intelligence techniques, neural networks have been considered significantly due to their ability to develop a non-linear function given large sets of data [2-4]. Support Vector Regression models have also been widely used to overcome the shortcomings of neural network models [5]. Other studies proposed a time-adaptive regression mode [6] and autoregressive-moving-average (ARMA) based models [7] for time series analysis. Wavelet based denoising algorithm [8] and adaptive method based on the multi-model partitioning algorithm [9] have also been used for short term load forecasting. Weather parameters have been taken into account for day-ahead load forecasting in [10] and an amalgamation of Fuzzy logic and artificial neural network model have been proposed in [11].

A Recurrent Neural Network (RNN) is a neural network with feedback [12]. The output of the network depends on the previously calculated outputs. The RNN is said to have a memory to store the calculations made so far and use them in the current computations. In addition to machine learning algorithms, another concept that is used to tackle the non-linear relationship between parameters is Fuzzy Logic. It allows us to apply a linguistic approach to theoretical and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

mathematical problems that are otherwise handled by the computer leading to binary solutions. It is governed by a set of Fuzzy rules according to which Fuzzy sets are formed. A hybrid approach of Fuzzy Logic and machine learning algorithms leads to an efficient model and accurate results. Neuro-Fuzzy system is homogeneous and represents neural networks having the advantageous characteristics of both Fuzzy logic and neural networks and hence it overcomes the drawbacks of their individual functionalities, resulting in higher accuracy. The non-linear relationship between parameters is handled well by the set of Fuzzy rules. The elements of Fuzzy sets are used as rather linear inputs in the neural network. This speeds up the training procedure of the neural network model and the accuracy attained is eventually higher.

In this paper the impact of integration of Fuzzy Logic with Neural Networks is observed and it is established that the hybrid model Fuzzy-RNN outperforms the other five models with the least Mean Absolute Percentage Error (MAPE) of 1.76. Hence, the MAPE value of Fuzzy-RNN lies in the range [1.64, 1.87].

This paper has been organized in six sections. The paper first briefly describes the neural network model used in Section II. The application of Fuzzy logic to devise a cause and effect relationship between weather parameters and electricity load demand has been depicted in Section III. Thereafter, the implementation of the proposed model including the selection of features, feature engineering, training and explanation of the performance metric is illustrated in Section IV. The performances of Fuzzy-RNN with other state of the art models have been compared for all the four seasons of a year in Section V. Finally, the paper is concluded in Section VI.

II. NEURAL NETWORKS

A neural network is an information processing system replicating the design and functioning of human brain. It has the ability to learn from example s, which gives it flexibility and makes it powerful. A neural network, also known as Artificial Neural Network (ANN), consists of highly interconnected processing elements called neurons. Each neuron is connected by other neurons by a connection link, which is associated with weights containing information about the link. Neural Network models can be classified on the basis of various criteria, like the network architecture, or the learning paradigm. The network architecture if defined by the arrangement of neurons to forms layers, and the connection pattern formed within and between layers. The most common network type is a single feed-forward network. In feed-forward network, input is fed into the network and transformed into an output, mapping raw data to categories. There's another type of architecture in which the input cycles through a loop. It's known as Recurrent Neural Network. A RNN takes input not just the current input example, but also what they have perceived previously in time. Another common type of neural network model is the Radial Basis Function Network. There are two types of RBFNs, one of them being Generic Regression Neural Network, which has a structure similar to radial basis network, with only slightly different second layer.

- *Recurrent Neural Network*

RNN are said to have an internal memory, since they store the output from the network to use it as an input in the next cycle of training. They have a feedback loop which ingests the calculated output in one cycle to the input for the next cycle. They're useful for sequential data since every neuron has its internal memory to store information about previous inputs. RNN is similar to feed forward Neural Network except that every layer has a recurrent connection with a tap delay associated with it. Because of the dependency of current input on the previous output, the network has an infinite dynamic response to time series input data.

III. FUZZY LOGIC MODEL

Unlike precise reasoning, Fuzzy logic is a multi-valued logic which handles reasoning approximately. The theory of Fuzzy logic and Fuzzy sets was created to allow computer to mimic the way humans think. In contrast to binary sets based on Boolean logic (either 0 or 1) which find solution to a particular set of problems, Fuzzy Logic has values ranging between 0 and 1. Since linguistic variables are used in Fuzzy logic, the degree with which a variable belongs to a Fuzzy set has to be managed by specific functions. The utility of Fuzzy sets lies in their ability to model ambiguous data and to provide suitable decisions.

Load demand has a non-linear dependence on parameters like temperature and humidity. In the proposed method Fuzzy logic has been used to handle this highly non-linear association between the weather elements and the electric load demand. Since fuzzy logic deals with cases that involve handling on a higher level than neural networks, it is used in pre-processing of the input data before feeding it to the neural network model. Unlike the internal layers of neural network model which is like a 'black box' to the user, the rules of Fuzzy inference model are declared by the user itself. The learning and optimizing capabilities of neural network are superior to Fuzzy logic. Combining the merits of both the models and disregarding their demerits leads to highly efficient results.

In the proposed model, Mamdani type Fuzzy Inference System (FIS) designing is implemented which calculates Fuzzy sets from IF-THEN Fuzzy rules [13]. These Fuzzy sets are further aggregated and defuzzified to output a crisp value as the final result of the process. The crisp inputs given to the Fuzzy inference module in the proposed method are:

i. change in dry bulb temperature: This is the difference between the forecasted dry bulb temperature during the hour for which load is to be predicted and previous day's same hour dry bulb temperature.
ii. change in dew point temperature*:* This is the difference between the forecasted dew point temperature during the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

hour for which load is to be predicted and previous day's same hour dew point temperature.

The dry bulb and dew point temperatures for the hour in which load demand needs to be predicted, are forecasted using a back propagation neural network which in turn acts as the inputs to the Fuzzy Logic Module.

The model quantifies the effect of dry bulb temperature and dew point temperature on load demand. There are five membership functions dividing the range of temperatures into five sections. Trial and error method was used to select the degree of overlap and the shape of membership functions. Trapezoid has been chosen as the shape. The only difference in membership function plots of dry bulb and dew point FIS lies in the ranges of input values. The difference is caused due to the fact that dew point temperature is always less than the dry bulb temperature.

TABLE I
FUZZY RULES SET

Previous Day Input	**Forecasted Day Input**				
	Very Cold	Cold	Comfo rtable	Hot	Very Hot
Very Cold (vc)	S	MN	LN	S	LP
Cold (co)	MP	S	MN	MP	LP
Comfortable (cf)	LP	MP	S	MP	LP
Hot (ho)	S	MN	MN	S	LP
Very Hot (vh)	LN	LN	LN	LN	S

25 IF-THEN rules are shown in Table I. The Fuzzy rule base was formed on the basis of prior knowledge and understanding as to how temperature affects the load demand. In Table I, LP, MP, S, MN and LN stand for large positive, medium positive, small, medium negative and large negative respectively. The required output of the module is described by the linguistic variables.

The output of the Fuzzy Logic Module is a weighted plot with weights lying between -1 and 1 which depicts the amount of change in load demand between the previous day and forecasted day. The Centroid defuzzification method converted the Fuzzy rules to one single output. The Fuzzy rules were implemented using the Fuzzy inference Max-Min composition.

IV. METHODOLOGY

The study and prediction was performed with a resolution of one hour on ISO New England dataset. The dataset is made freely accessible on the ISO New England Integration (ISO-NE) official website. Fuzzy-RNN's proposed methods were contrasted with independently-used RNN[14] and GRNN models and other models, including Fuzzy- Artificial Neural Network (ANN)[15], ANN[16], Support Vector Machine (SVM)[17], and Generic Regression Neural Network (GRNN)[18].

The error metric Mean absolute percentage error has been used to validate the forecasted load. Publicly accessible is the ISO New England dataset which allows researchers to compare their models with the models proposed in this paper. This allows for fair comparison and assessment of different methodologies in load forecasting. The implementation has been elucidated in the succeeding subsections.

A. Dataset and Features

The ISO New England dataset used comprises of hourly load demand for a period of two years from 2013 to 2014. Hence the dataset contains 17520 observations which are further divided into training and testing datasets. The first three weeks of each month have been included in the training set and the last week of each month was incorporated in the test set. As the purpose is short term load forecasting, the test data was used to predict the load demand of the next day for the same hour. For each hour, the dataset contains information on the average load demand, dry bulb temperature in degree Fahrenheit, dew point temperature in degree Fahrenheit and the day ahead local marginal price.

Dry bulb temperature and dew point temperature were inputs to the Fuzzy Logic model which computed the change in load demand in the next day for the same hour. This change in demand was included as a feature in the neural network model. Two new features were introduced in the dataset in order to improve the computation and accuracy of the proposed models. The day of the load being considered was introduced in numerical form (Sunday=1, Monday=2, Tuesday=3…) as it is highly likely that the load demands of the same days of a week are correlated owing to the routine lifestyle of human beings. Another binary feature presented is based on the pattern of higher social activities on the weekdays compared to weekends. The load demand is generally higher on the weekdays than on the weekends. Hence in this feature, the weekends are set to 1 and the weekdays are set to 0 centered on the concept of one hot encoding in applied machine learning.

Without Fuzzy Logic model, the maximum and minimum temperature of the day were used as features instead of change in load demand mentioned in the proposed models.

B. Method

The dataset contained a few missing values of load demand. During preprocessing, these missing values were completed by interpolating between neighboring values. The features were normalized by subtracting the mean value from each

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

TABLE II
SEASON-WISE MAPE FOR EACH MODEL

Season	Fuzzy-RNN	RNN[14]	Fuzzy-ANN[17]	ANN[16]	SVM[17]	GRNN[18]
Spring Week	1.71	2.82	2.13	3.67	2.69	3.55
Summer Week	1.73	3.67	2.68	3.57	3.31	5.01
Fall Week	1.75	2.03	2.24	2.69	1.57	2.80
Winter Week	1.86	2.85	2.45	3.39	2.90	4.30
Average	**1.76**	**2.84**	**2.38**	**3.33**	**2.62**	**3.92**

observation and dividing by its standard deviation from the mean.

The Fuzzy Logic and neural network models were developed and were further tested using the test set described earlier. Fig. 1 shows the plot of the predicted load vs. the actual load for 48 hours starting from the first hour of 21/02/2013. For fair assessment, the plots of the same period of time have been shown using all the four models. As can be observed from the figures, the Fuzzy-RNN model depicts the best performance as the curve of the predicted load fits the actual load curve almost perfectly.

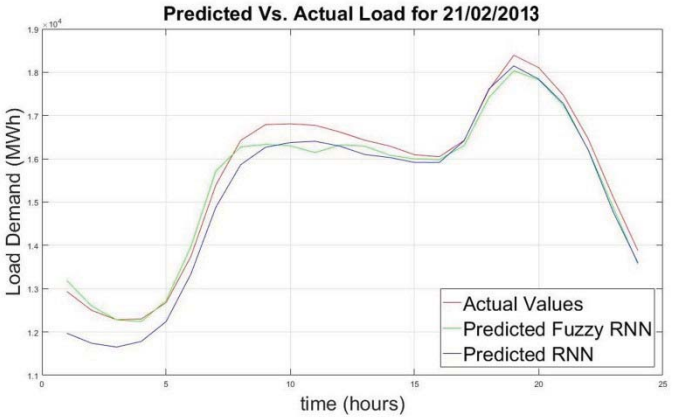


Fig. 1. Predicted vs. actual load values for Fuzzy-RNN

C. *Block Diagram*

Fig. 2 illustrates the block diagram of the proposed neuro-Fuzzy model for short term load forecasting. In the two level model, the first level depicts the Fuzzy logic which maps the weather parameters with load demand as described in the earlier section. In the second level, the input features are provided to the neural network model for training and hence predicting the day ahead load demand in an electricity market.

V. RESULTS

The neural network models trained were optimized using random grid search for hyper parameter tuning. Concretely, the Recurrent neural network model was found to give minimum cross validation error when trained using three hidden layers consisting of 15 neurons each. Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) algorithm [19]

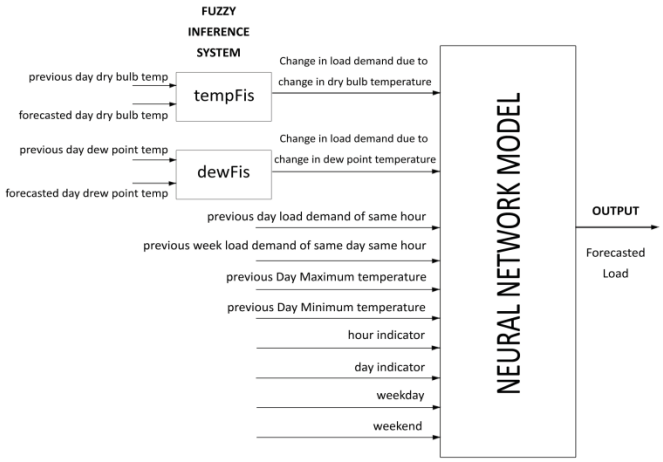


Fig. 2. Block Diagram of Hybrid Fuzzy-NN Model

was used to train the RNN model. It is a second order optimization method that uses rank-one updates. The updates are specified by evaluations of gradient g based on [20]. The algorithm reached convergence on 200 iterations.

The performance of the six models implemented to forecast load demand of the next day is now compared. For fair assessment, four different weeks of all the four seasons of the year 2013 have been chosen from the test set. As shown in Table II, the Mean Absolute Percentage Error (MAPE) of the four weeks is evaluated using the neuro-Fuzzy models proposed. The observed hourly MAPE values of one month are plotted and shown in Fig. 3.

It has been observed that using Fuzzy Logic approach to handle the parameters linguistically and then passing the outputs of Fuzzy Inference System as input to the Neural Network, majorly improves the efficiency of the whole model. Combining the merits of both approaches (Fuzzy Logic & Neural Network) progresses the performance of the load forecasting system.

Fig. 4 depicts hourly Absolute Percentage Error (APE) values for a randomly chosen day 21/02/2013 from the test dataset. Specifically Fig. 4 compares hourly APE values of RNN and Fuzzy-RNN.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

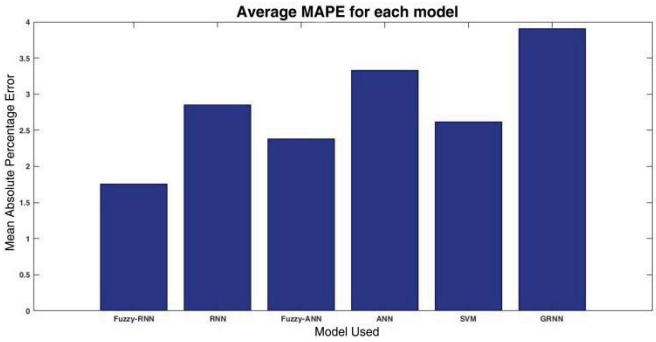


Fig. 3. Load error comparison of each model

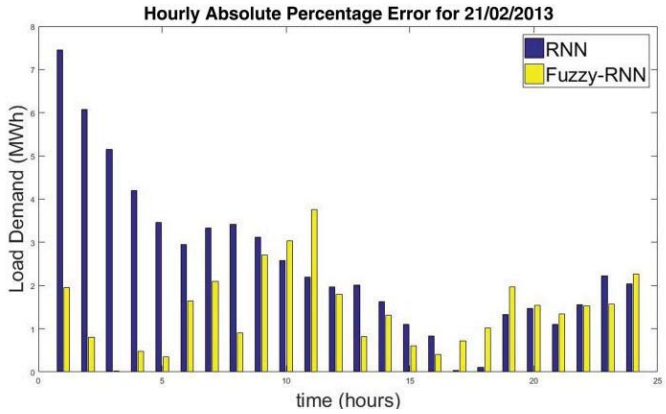


Fig. 4. Hourly Absolute Percentage Error plot of a day for Fuzzy-RNN

The plots show high irregularities in some regions for the APE values of RNN implementation. These irregularities have been considerably reduced in the proposed hybrid model of Fuzzy-RNN. Combining the merits of both approaches progresses the performance of the load forecasting system.

VI. CONCLUSION

In this paper, an efficient algorithm is based on hybrid approach of Fuzzy Logic & Neural Networks is proposed. Six different models for load forecasting are implemented and compared. The Fuzzy-RNN model outperforms the other models with same parameters and dataset. The load demand of next day for the same hour is forecasted competently using the proposed algorithm with 1.76 MAPE on test dataset making it the most efficient among the four models implemented.

REFERENCES

[1] X. Sun et al., "An Efficient Approach to Short-Term Load Forecasting at the Distribution Level," in IEEE Transactions on Power Systems, vol. 31, no. 4, pp. 2526-2537, July 2016.

[2] D. Park, M. El-Sharkawi, R. J. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," IEEE Transactions on Power Systems, vol. 6, no. 2, pp. 442–449, May 1991.

[3] A. Piras, A. Germond, B. Buchenel, K. Imhof, and Y. Jaccard, "Heterogeneous artificial neural network for short term electrical load forecasting," IEEE Transanctions on Power Systems, vol. 11, no. 1, pp. 397–402, Feb. 1996.

[4] M. Daneshdoost, M. Lotfalian, G. Baumroonggit, and J. P. Ngoy, "Neural network with Fuzzy set-based classification for short-term load forecasting," IEEE Transactions on Power Systems, vol. 13, no. 4, pp. 1386-1391, Nov. 1998.

[5] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," IEEE Transactions on Power Systems, vol. 19, no. 4, pp. 1821–1830, Nov. 2004.

[6] T. Jensson, P. Pinson, H. Madsen, and H. A. Nielsen, "Predictive densities for day-ahead electricity prices using time-adaptive quantile regression," Energies, vol. 7, no. 9, pp. 5523–5547, 2014.

[7] H. Liu and J. Shi, "Applying ARMA–GARCH approaches to forecasting short-term electricity prices," Energy Economics, vol. 37, pp. 152–166, May 2013.

[8] Ekonomou L., Christodoulou C.A., Mladenov V., A short-term load forecasting method using artificial neural networks and wavelet analysis, International Journal of Power Systems, Vol. 1, pp. 64-68, 2016.

[9] Pappas S.Sp., Ekonomou L., Moussas V.C., Karampelas P., Katsikas S.K., Adaptive load forecasting of the Hellenic electric grid, Journal of Zhejiang University SCIENCE A, Vol. 9, No. 12, pp. 1724-1730, 2008.

[10] S.Fan, K. Methaprayoon, and W. J. Lee, "Short-term multi-region load forecasting based on weather and load diversity analysis," in Proc. 39th North Amer. Power Symp., pp. 562–567, 2007.

[11] A. Khotanzad, E. Zhou, and H. Elragal, "A neuro-Fuzzy approach to short-term load forecasting in a price-sensitive environment," IEEE Transactions on Power Systems, vol. 17, no. 4, pp. 1273–1282, Nov. 2002.

[12] O. Olurotimi, "Recurrent neural network training with feedforward complexity," IEEE Transactions on Neural Networks, vol. 5, pp. 185-197, March 1994.

[13] H.-Jürgen, Zimmermann, Fuzzy Set Theory-and Its Applications. 4th ed., Springer Netherlands, Netherlands, 2001.

[14] J. Vermaak, E. C. Botha, "Recurrent Neural Networks for Short-Term Load Forecasting", IEEE Transactions on Power Systems, vol. 13, no. 1, pp. 126-132, February 1998.

[15] M. W. Xiao, B. X. Min, M. L. Shun, "Short-Term Load Forecasting With Artificial Neural Network And Fuzzy Logic", Proceedings of Power System Technology, 2002, PowerCon 2002, October 2002.

[16] S. Singh, S. Hussain, M. A. Bazaz, "Short Term Load Forecasting Using Artificial Neural Network", IEEE 4th International Conference on Image Information Processing (ICIIP), pp. 159-163, 2017.

[17] N. Ye, Y. Liu, Y. Wang, " Short-Term Power Load Forecasting Based on SVM", Proceedings of 2010 Conference on Dependable Computing (CDC'2010), pp. 47-51, 2010.

[18] Smys, S., Robert Bestak, and Joy Iong-Zong Chen. "Special issue on evolutionary computing and intelligent sustainable systems." (2019): 1-1.

[19] N. M. Nawi, M. R. Ransing, R S. Ransing, "An Improved Learning Algorithm Based on The Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method For Back Propagation Neural Networks", Proceedings of Sixth International Conference on Intelligent Systems Design and Applications, 2006 (ISDA '06), October 2006.

[20] P.J. Werbos, "Backpropagation through time: what it does and how to do it'', Proceedings of the IEEE, vol. 78, no. 10, pp. 1550-1560, 1990.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:07 UTC from IEEE Xplore. Restrictions apply.

Physics of Plasmas

The effect of dust grains on the Weibel instability in presence of large amplitude electrostatic waves

Cite as: Phys. Plasmas **27**, 043702 (2020); https://doi.org/10.1063/1.5133756
Submitted: 30 October 2019 . Accepted: 05 March 2020 . Published Online: 01 April 2020

Ajay Gahlot, Suresh C. Sharma, and Jyotsna Sharma


View Online

Export Citation

CrossMark

ARTICLES YOU MAY BE INTERESTED IN

A study of the propagation of a solitary wave along the magnetic field in a cold collision-free plasma
Physics of Plasmas **27**, 042102 (2020); https://doi.org/10.1063/1.5142704

A study of SCL sheath in presence of two dust species in a photoemissive environment
Physics of Plasmas **27**, 043701 (2020); https://doi.org/10.1063/1.5130551

Interaction of electron acoustic waves in the presence of superthermal electrons in terrestrial magnetosphere
Physics of Plasmas **27**, 042105 (2020); https://doi.org/10.1063/1.5143400



AIP Publishing

© 2020 Author(s).

Physics of Plasmas ARTICLE scitation.org/journal/php

The effect of dust grains on the Weibel instability in presence of large amplitude electrostatic waves

Cite as: Phys. Plasmas **27**, 043702 (2020); doi: 10.1063/1.5133756
Submitted: 30 October 2019 · Accepted: 5 March 2020 ·
Published Online: 1 April 2020

View Online Export Citation CrossMark

Ajay Gahlot,[1] Suresh C. Sharma,[2,a)] and Jyotsna Sharma[3]

AFFILIATIONS

[1]Department of Applied Sciences, MSIT, C-4, Janakpuri, Delhi 110058, India
[2]Department of Applied Physics, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India
[3]Department of Physics, Amity School of Applied Sciences, Amity University Haryana, Gurugram 122413, India

[a)]Author to whom correspondence should be addressed: suresh321sharma@gmail.com

ABSTRACT

Theoretical investigations on the effect of dust charge fluctuations on Weibel instability in a plasma, with an electron beam in the presence of large amplitude electrostatic Langmuir waves, have been performed. It is observed that the dust charge fluctuations enhance the growth rate of electromagnetic (EM) perturbations. An increase in the electron to ion temperature ratio amplifies the EM perturbations, leading to an enhanced magnetic field due to Weibel instability. It has been found that the growth rate of Weibel instability strongly depends on the dust grain size for different values of ion to electron density ratio δ_m. Moreover, the electron to ion temperature ratio also destabilizes the Weibel instability.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5133756

I. INTRODUCTION

Experiments involving laser produced plasmas support a stream of drifting hot electrons,[1] which can be modeled as an electron beam propagating through plasmas. In such cases, amplification of electromagnetic (EM) perturbations leads to the generation of a magnetic field on the account of Weibel instability. In recent years, research involving the parametric effects of electrostatic[2–6] and EM[7,8] waves having large amplitude is of much interest due to its profound applications in space plasmas, fusion devices,[9] and laboratory plasmas[10–12] as well as in the phenomenon where plasma interacts with the laser.[13]

Electrostatic waves in dusty plasmas, which are characterized by a low concentration of suspended dust grains, have been exploited prominently in recent years.[14–36] Chow *et al.*[16,17] modeled electrostatic ion cyclotron (EIC) instability in complex plasma using Vlasov theory, and the results of their study were in accordance with the investigations of Barkan *et al.*[18] Ma and Yu[27] showed that the exchange of energy with the dust charging process results in the damping of ion acoustic waves. Recently, Gahlot[36] investigated the modulational instability of lower hybrid waves in dusty plasma.

It is well known that the presence of dust particles has an impact on the parametric processes in magnetized[37] and unmagnetized plasmas.[38–40] Annou and Tripathi[41] have investigated Weibel instability in the presence of large amplitude Langmuir waves, which can be generated due to parametric processes like Raman scattering or by the laser through linear mode conversion.

The present paper investigates the effect of dust grains on the Weibel instability in the presence of large amplitude electrostatic Langmuir waves. Here, the EM perturbation $(\omega_m, \vec{k}_m)$ pairs with a large amplitude electrostatic Langmuir pump wave $(\omega_0, \vec{k}_0)$ to produce two Langmuir sidebands $(\omega_{1,2} = \omega_m \mp \omega_0, \vec{k}_{1,2} = \vec{k}_m \mp \vec{k}_0)$. These sidebands generate a nonlinear current at $(\omega_m, \vec{k}_m)$ on interacting with the pump, thus amplifying the original EM perturbations. Section II shows the complete instability analysis using fluid treatment, while Sec. III presents the results and discussion part. Conclusions are stated in Sec. IV.

II. INSTABILITY ANALYSIS

We consider a plasma consisting of electrons, ions, and dust grains having equilibrium charge, density, mass, and temperature as $(-e, n_{pe}, m_e, T_e)$, (e, n_{pi}, m_i, T_i), and $(-Q_{pd}, n_{pd}, m_d, T_d)$, respectively. An electron beam of velocity v_{be} and density n_{be} propagates through plasma along the z-direction. The potential of large amplitude Langmuir waves existing in plasma is of the form

$$\phi_0 = \phi_0(z) \exp\left[-i(\omega_0 t - k_0 z)\right],$$

Published under license by AIP Publishing

where $\omega_0^2 = \omega_{pe}^2 + k_0^2 v_{th}^2$ and $\omega_{pe}\left(=\sqrt{\frac{4\pi n_{pe}e^2}{m_e}}\right)$ and $v_{th}\left(=\sqrt{\frac{T_e}{m_e}}\right)$ are the electron plasma frequency and electron thermal velocity, respectively.

The linear response of beam and plasma electrons to the Langmuir pump wave is obtained using the equation of motion $m_e\frac{d\vec{v}}{dt} = -e\vec{E}$ and equation of continuity[43,44] $\frac{\partial n}{\partial t} + \nabla.(n\vec{v}) = 0$.

The perturbed velocity and density of beam electrons are given as

$$\vec{v}_{be0} = \frac{-ek_0\phi_0}{m_e[\omega_0 - k_0 v_{be}]}\hat{z}, \tag{1}$$

$$n_{be0} = \frac{-ek_0^2 n_{be}\phi_0}{m_e[\omega_0 - k_0 v_{be}]^2}. \tag{2}$$

Similarly, for plasma electrons, we get

$$\vec{v}_{pe0} = \frac{-ek_0\phi_0}{m_e\omega_0}\hat{z}, \tag{3}$$

$$n_{pe0} = \frac{-ek_0^2 n_{pe}\phi_0}{m_e\omega_0^2}. \tag{4}$$

This system is perturbed by an EM perturbation

$$\vec{E} = \vec{E}_m \exp\left[-i(\omega_m t - \vec{k}_m.\vec{x})\right].$$

The magnetic field $\vec{B}$ is polarized along the y-direction. The nonlinear coupling between the Langmuir pump wave and EM perturbation generates two Langmuir sidebands having potentials as $\phi_1 = \phi_1(r)\exp[-i(\omega_1 t - \vec{k}_1.\vec{r})]$ and $\phi_2 = \phi_2(r)\exp[-i(\omega_2 t - \vec{k}_2.\vec{r})]$, respectively, where $\omega_{1,2} = \omega_m \mp \omega_0$, and $\vec{k}_{1,2} = k_m\hat{x} \mp k_0\hat{z}$.

The linear response of beam electrons to ϕ_1 and ϕ_2 is given as

$$\vec{v}_{bej} = \frac{-ek_j\phi_j}{m_e\left[\omega_j - k_j v_{be}\right]}\hat{z} \tag{5}$$

and

$$n_{bej} = \frac{-ek_j^2 n_{be}\phi_j}{m_e\left[\omega_j - k_j v_{be}\right]^2} = \frac{k_j^2\chi_{bej}\phi_j}{4\pi e}, \tag{6}$$

where j = 1, 2. $\chi_{bej} = \frac{-\omega_{be}^2}{[\omega_j - k_j v_{be}]^2}$ is the beam electron susceptibility and $\omega_{be}^2 = \frac{4\pi n_{be}e^2}{m_e}$.

Similarly, the linear plasma electron response to ϕ_1 and ϕ_2 is given by

$$\vec{v}_{pej} = \frac{-ek_j\phi_j}{m_e\omega_j}\hat{z} \tag{7}$$

and

$$n_{pej} = \frac{-ek_j^2 n_{pe}\phi_j}{m_e\omega_j^2} = \frac{k_j^2\chi_{pej}\phi_j}{4\pi e}, \tag{8}$$

where $\chi_{pej} = \frac{-\omega_{pe}^2}{\omega_j^2}$ is the plasma electron susceptibility.

The magnetic field of EM perturbation combines with the oscillatory velocity of plasma electrons ($\vec{v}_{pe0}$) provided by Langmuir pump waves to generate ponderomotive forces $\vec{F}_{P1,2}$ at $(\omega_{1,2}, \vec{k}_{1,2})$ as

$$\vec{F}_{p1} = \frac{-e}{2C}\left(\vec{v}^*_{pe0} \times \vec{B}\right) = ie\vec{k}_m\phi_{p1},$$

where $\vec{v}^*_{pe0}$ denotes the complex conjugate of $\vec{v}_{pe0}$,

$$\vec{F}_{p2} = \frac{-e}{2C}\left(\vec{v}_{pe0} \times \vec{B}\right) = ie\vec{k}_m\phi_{p2},$$

where $\phi_{p1} = -\frac{\vec{v}^*_{pe0}.\vec{E}}{2i\omega_m}$, $\phi_{p2} = -\frac{\vec{v}_{pe0}.\vec{E}}{2i\omega_m}$, $\vec{B} = \frac{C(\vec{k}_m \times \vec{E})}{\omega_m}$, and $C = 3 \times 10^8$ m/s.

The nonlinear plasma electron responses to ponderomotive potential ϕ_{p1} and ϕ_{p2} are given by

$$n^{NL}_{pej} = \frac{k_m^2\chi_{pej}\phi_{pj}}{4\pi e}, \tag{9}$$

where j = 1, 2.

Similarly, for beam electrons, we obtain

$$n^{NL}_{bej} = \frac{k_m^2\chi_{bej}\phi_{bj}}{4\pi e},$$

where

$$\phi_{b1} = -\frac{\vec{v}^*_{be0}.\vec{E}}{2i\omega_m} \text{ and } \phi_{b2} = -\frac{\vec{v}_{be0}.\vec{E}}{2i\omega_m}.$$

Following Annou and Tripathi,[41] the linear ($\vec{J}^L$) and nonlinear current density($\vec{J}^{NL}$) at $(\omega_m, \vec{k}_m)$ is given by

$$\vec{J}^L = -\frac{e^2E}{im_e\omega_m}\left[n_{pe} + n_{be} + \frac{n_{be}k_m^2v_{be}^2}{\omega_m^2}\right]\hat{z} - \frac{n_{be}e^2k_m v_{be}E}{im_e\omega_m^2}\hat{x},$$

$$\vec{J}^{NL} = -\frac{en_{pe1}\vec{v}_{pe0}}{2} - \frac{en_{pe2}\vec{v}^*_{pe0}}{2} - \frac{en_{be1}\vec{v}_{be0}}{2} - \frac{en_{be2}\vec{v}^*_{be0}}{2}, \tag{10}$$

$$= -\frac{k_1^2}{8\pi}\left[\chi_{pe1}\vec{v}_{pe0} + \chi_{be1}\vec{v}_{be0}\right]\phi_1 - \frac{k_2^2}{8\pi}\left[\chi_{pe2}\vec{v}^*_{pe0} + \chi_{be2}\vec{v}^*_{be0}\right]\phi_2. \tag{11}$$

Dust charge fluctuations are obtained by following Jana *et al.*,[15] i.e.,

$$Q_{pd1} = \frac{|I_{pe}|}{i(\omega_m + i\eta_m)}\left(\frac{n_{i1}}{n_{pi}} - \frac{n_{e1}}{n_{pe}}\right), \tag{12}$$

where n_{e1} and n_{i1} are the perturbed electron and ion density, respectively. η_m is the dust charging rate given as $\eta_m = 0.79a_0\left(\frac{\omega_{pi}}{\lambda_{di}}\right)\left(\frac{1}{\delta_m}\right)\left(\frac{m_i}{m_e}\frac{T_i}{T_e}\right)^{\frac{1}{2}} \sim 10^{-2}\omega_{pe}\left(\frac{a_0}{\lambda_{de}}\right)\frac{1}{\delta_m}$, where λ_{de} and λ_{di} are the electron and ion Debye length, respectively, while "a_0" is the dust grain size.

$\omega_{pi}\left(=\sqrt{\frac{4\pi n_{pi}e^2}{m_i}}\right)$ denotes the ion plasma frequency. In Eq. (12), we assume that the dust charging time (η_m^{-1}) is nearly equal to wave period (ω_m^{-1}).

In equilibrium, the quasi-neutrality condition gives

$$en_{pe} + en_{be} - en_{pi} + Q_{pd}n_{pd} = 0; \tag{13}$$

as $n_{pe} \gg n_{be}$, we get

$$\frac{n_{pd}}{n_{pe}} = (\delta_m - 1)\frac{e}{Q_{pd}}, \quad \text{where } \delta_m = \frac{n_{pi}}{n_{pe}}. \tag{14}$$

Published under license by AIP Publishing

Substituting the perturbed densities in Poisson's equation, we obtain mode structure equations for sidebands ϕ_1 and ϕ_2 in the presence of dust,

$$\phi_j = \frac{-k_m^2}{k_j^2 \varepsilon_{jd}}\left[1+\left(\frac{i\beta_m}{\omega_j + i\eta_m}\right)\right]\left[\chi_{bej}\phi_{bj} + \chi_{pej}\phi_{pj}\right], \tag{15}$$

where

$$\varepsilon_{jd} = 1 + \left[1+\left(\frac{i\beta_m}{\omega_j + i\eta_m}\right)\right]\left[\chi_{bej} + \chi_{pej}\right] \text{and j} = 1,\ 2. \tag{16}$$

Here, we assume $\frac{i\beta_m}{\omega_j + i\eta_m} \ll 1$ in Eqs. (15) and (16).

Following Annou and Tripathi,[41] we obtain perturbed density of beam electrons (n_{bem}) and plasma ions (n_{im}) due to EM waves at $(\omega_m, \vec{k}_m)$ as

$$n_{bem} = \frac{n_{be} e k_m^2 v_{be} E}{m_e i\omega_m^3} \text{ and } n_{im} = 0.$$

The wave equation for EM waves is given as

$$\nabla(\nabla.\vec{E}) - \nabla^2\vec{E} = \frac{4\pi i\omega_m}{C^2}\left[\vec{J}^L + \vec{J}^{NL}\right] + \frac{\omega_m^2}{C^2}\vec{E}. \tag{17}$$

Now,

$$\nabla.\vec{E} = -4\pi n_{pd} Q_{pd1} = \frac{-4\pi\beta_m n_{be} e^2 k_m^2 v_{be} E}{(\omega_m + i\eta_m) m_e \omega_m^3}, \tag{18}$$

where $\beta_m = \frac{n_{pd}|I_{pe}|}{e n_{pe}} = 0.397\left(1-\frac{1}{\delta_m}\right)\left(\frac{a_0}{V_{ith}}\right)\left(\frac{m_i}{m_e}\right)\omega_{pi}^2$ is the coupling parameter.

Now, using Eqs. (10), (11), and (18) in wave equation (17), we get

$$\left[k_m^2 - \frac{\omega_m^2}{C^2} + \frac{\omega_{pe}^2}{C^2} + \frac{\omega_{be}^2}{C^2}\left(1+\frac{k_m^2 v_{be}^2}{\omega_m^2}\right) - k_m^2 G_D\right]\vec{E}$$
$$= \frac{-i\omega_m}{2C^2}\left[\left(\chi_{pe1}\vec{v}_{pe0} + \chi_{be1}\vec{v}_{be0}\right)k_1^2\phi_1 + \left(\chi_{pe2}\vec{v}_{pe0}^* + \chi_{be2}\vec{v}_{be0}^*\right)k_2^2\phi_2\right], \tag{19}$$

where

$$G_D = \frac{i4\pi\beta_m n_{be} e^2 v_{be} \varepsilon_{eff}^{\frac{1}{3}}}{(\omega_m + i\eta_m) m_e \omega_m^2 C},\ \varepsilon_{eff} = 1 - \frac{\omega_{pe}^2}{\omega_m^2}.$$

First substituting the values of ϕ_1 and ϕ_2 from Eq. (15) and then substituting the values of $\phi_{p1}, \phi_{p2}, \phi_{b1}$ & ϕ_{b2} in Eq. (19), we get

$$\left[k_m^2 - \frac{\omega_m^2}{C^2} + \frac{\omega_{pe}^2}{C^2} + \frac{\omega_{be}^2}{C^2}\left(1+\frac{k_m^2 v_{be}^2}{\omega_m^2}\right) - k_m^2 G_D\right]$$
$$= \frac{-k_m^2}{4C^2\varepsilon_{1d}}\left[\chi_{pe1}v_{pe0} + \chi_{be1}v_{be0}\right]\left[\chi_{be1}v_{be0}^* + \chi_{pe1}v_{pe0}^*\right]$$
$$- \frac{k_m^2}{4C^2\varepsilon_{2d}}\left[\chi_{pe2}v_{pe0}^* + \chi_{be2}v_{be0}^*\right]\left[\chi_{be2}v_{be0} + \chi_{pe2}v_{pe0}\right].$$

Using Eqs. (1) and (3), we get

$$\left[k_m^2 - \frac{\omega_m^2}{C^2} + \frac{\omega_{pe}^2}{C^2} + \frac{\omega_{be}^2}{C^2}\left(1+\frac{k_m^2 v_{be}^2}{\omega_m^2}\right)\right] = k_m^2 G_D - \frac{k_m^2 |v_{pe0}|^2}{4C^2}\xi(\omega_m),$$

where

$$\xi(\omega_m) = \frac{1}{\varepsilon_{1d}}\left[\chi_{pe1} + \frac{\chi_{be1}}{1-\frac{k_0 v_{be}}{\omega_0}}\right]^2 + \frac{1}{\varepsilon_{2d}}\left[\chi_{pe2} + \frac{\chi_{be2}}{1-\frac{k_0 v_{be}}{\omega_0}}\right]^2.$$

For larger values of k_m, the above equation can be rewritten as

$$1 + \frac{\omega_{be}^2 v_{be}^2}{C^2\omega_m^2} = \frac{\omega_0\mu|v_{pe0}|^2}{4C^2(\omega_m^2 - \mu^2)} + G_D, \tag{20}$$

where

$$\mu = \omega_0 - \sqrt{\omega_{pe}^2 + k_1^2 v_{th}^2}. \tag{21}$$

The roots of the above quadratic equation are given as

$$\omega_m^2 = \frac{\left[\mu^2 + \frac{|v_{pe0}|^2\omega_0\mu}{4C^2} - \frac{\omega_{be}^2 v_{be}^2}{C^2} - G_D\mu^2\right] \pm \left\{\left[-\mu^2 - \frac{|v_{pe0}|^2\omega_0\mu}{4C^2} + \frac{\omega_{be}^2 v_{be}^2}{C^2} + G_D\mu^2\right]^2 - 4(1-G_D)\left(\frac{-\omega_{be}^2 v_{be}^2\mu^2}{C^2}\right)\right\}}{2(1-G_D)}.$$

When $|v_{pe0}| = 0$, i.e., there is no pump wave or $\mu = 0$, we get

$$2(1-G_D)\omega_m^2 = -\frac{\omega_{be}^2 v_{be}^2}{C^2} \pm \frac{\omega_{be}^2 v_{be}^2}{C^2}.$$

Taking negative sign in the above equation and $\omega_m = \omega_r + i\gamma_m$ with $|\gamma_m| \ll \omega_r$, we obtain the growth rate of EM perturbation as

$$\gamma_m = \frac{\omega_{be} v_{be}}{C\sqrt{1-G_D}}. \tag{22}$$

We obtain the dispersion relation of Annou and Tripathi[41] (cf. p. 195), in the absence of dust grains by substituting $\delta_m = 1$ and $\beta_m = 0$.

III. RESULTS AND DISCUSSION

We have theoretically investigated the impact of dust charge fluctuations (DCF), size of charged dust grains a_0, and different values of $\left(\frac{T_e}{T_i}\right)$ ratio on Weibel instability in a complex plasma in the presence of electrostatic Langmuir waves. The parameters used in the study are: $n_{pi} = 2.0 \times 10^{10}$ cm^{-3}, $n_{be} = 1.0 \times 10^7$ cm^{-3}, $n_{pd} = 2.0 \times 10^4$ cm^{-3},

Published under license by AIP Publishing

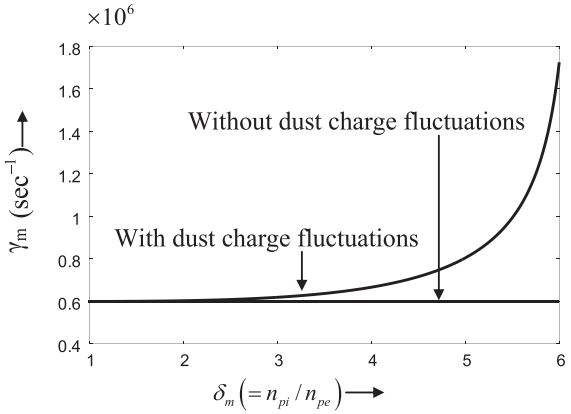


FIG. 1. Growth rate γ_m (s^{-1}) of the EM perturbation as a function of the density ratio of negatively charged dust grains $\delta_m(=n_{pi}/n_{pe})$ with and without dust charge fluctuations. The parameters are given in the text.

$T_e = 0.4$ eV, $T_i = 0.1$ eV, $m_i/m_e \approx 7.16 \times 10^4$ (potassium), $a_0 = 5\ \mu$m, and $v_{be} = 1.0 \times 10^8$ cm/s.

$\delta_m(=n_{pi}/n_{pe})$ is varied from 1 to 6. In Fig. 1, the growth rate of EM perturbation $\gamma_m(s^{-1})$ is plotted as a function of $\delta_m(=n_{pi}/n_{pe})$, both in the presence as well as the absence of DCF. γ_m shows an increase by a factor of ∼1.27 when δ_m changes from 1 to 6 in the presence of DCF (i.e., η_m is finite). As n_{pi} is constant, the increase in the value of δ_m decreases n_{pe} which indicates that more electrons get attached to the dust grains. The impact of DCF on γ_m is similar to the experimental results of Barkan *et al.*[18] Similarly, in the absence of DCF (i.e., $Q_{pd_1} = 0$ when $\eta_m \to \infty$), $\vec{\nabla}.\vec{E} = -4\pi n_{pd} Q_{pd1} = 0$ and hence γ_m remains constant for all the values of δ_m. The frequencies of the waves involved in the present channel (Langmuir and EM) are relatively high and the dust component is not able to respond to these frequencies.

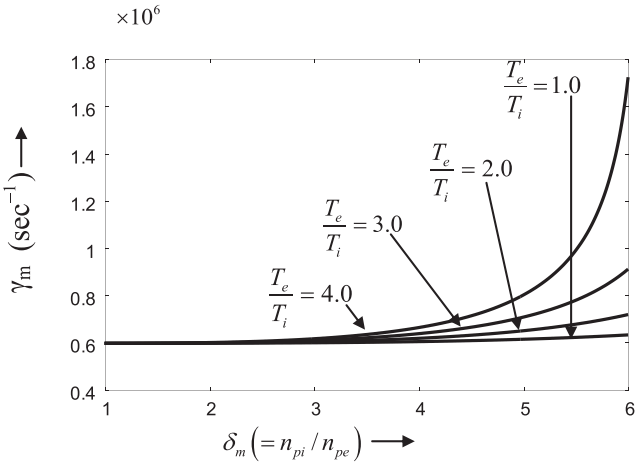


FIG. 2. Growth rate γ_m (s^{-1}) of the EM perturbation as a function of the density ratio of negatively charged dust grains $\delta_m(=n_{pi}/n_{pe})$ for the same parameters as in Fig. 1 and for different values of electron to ion temperature ratios.

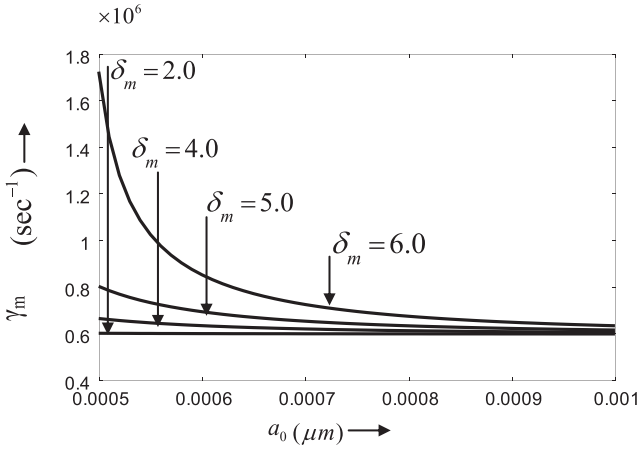


FIG. 3. Growth rate γ_m (s^{-1}) of the EM perturbation as a function of dust grain size $a_0(\mu m)$ for the same parameters as in Fig. 1 and for different values of δ_m.

Figure 2 shows the variation of $\gamma_m(s^{-1})$ with δ_m for various values of $\frac{T_e}{T_i}$ ratio.

However, the EM perturbation grows exponentially for higher values of δ_m (≥3), for all values of $\frac{T_e}{T_i}$, but this increase is more pronounced at higher values of electron to ion temperature ratio. $\gamma_m(s^{-1})$ is increased by a factor of ∼2.72 as electron temperature increases from 0.1 to 0.4 eV corresponding to $\delta_m = 6$. The increase in electron temperature results in the increased thermal velocity and hence more electrons get attached to the dust grains. Moreover, the enhanced dust grain charge results in an increased growth rate of EM perturbations.

The variation of $\gamma_m(s^{-1})$ with dust grain size a_0 (in μm) for different values of δ_m is depicted in Fig. 3. It can be observed that increased dust grain size results in exponential wave damping for all values of δ_m. This can be explained on the basis of the screening effect as observed by Wang and Bhattacharjee.[42] The increase in the dust grain size reduces the intergrain spacing, which in turn makes the screening of dust grain charge over other dust grains significant and results in decreased potential of dust grains. The reduced dust grain charge tends to decrease the growth rate with increasing dust grain size for a particular value of δ_m. Equation (14) shows that dust grain charge Q_{pd} increases with the increase in δ_m at constant ion-dust densities. Thus, the decrease in grain potential is more pronounced for larger values of δ_m with increasing values of a_0, which in turn results in sharp damping of EM perturbations at higher values of δ_m. The growth rate shows no significant change beyond $a_0 = 10\ \mu$m for higher values of δ_m.

IV. CONCLUSION

In the present work, we have given a detailed fluid description to investigate the impact of dust grains on Weibel instability that may arise in laser fusion plasma on account of anisotropic velocity distribution (leading to hot electron beams) in the presence of Langmuir waves. The destabilization effect of the DCF and ion–electron temperature ratio on the growth of the magnetic field generated due to Weibel instability has been explicitly shown. The enhancement in the magnetic field is significant at shorter wavelengths of EM radiations, i.e., for larger values of k_m. The increased size of dust grains (from 5 μm to 10 μm) has a stabilizing effect on the growth of unstable mode

Published under license by AIP Publishing

instability, which strongly depends upon δ_m. This study can be extended further and may be of significant use in analyzing the EM fluctuations due to velocity or temperature anisotropy in solar plasma that is hot and almost collisionless. This can probably help in understanding the production of magnetic fields in laser induced plasma experiments due to Weibel instability. The generated magnetic field can also affect EM propagation and transport of electron energy in fusion schemes involving inertial confinement.

REFERENCES

[1]W. P. Leemans, C. E. Clayton, W. B. Mori, K. A. Marsh, P. K. Kaw, A. Dyson, C. Joshi, and J. M. Wallace, Phys. Rev. A **46**, 1091 (1992).
[2]N. S. Wolf, R. Majeski, H. Lashinky, V. K. Tripathi, and C. S. Liu, Phys. Rev. Lett. **45**, 799 (1980).
[3]C. S. Liu, V. K. Tripathi, V. S. Chan, and V. Stefan, Phys. Fluids **27**, 1709 (1984).
[4]K. Akimoto, Phys. Fluids **B1**, 1998 (1989).
[5]C. S. Liu, V. S. Chan, D. K. Bhandari, and R. W. Harvey, Phys. Fluids **48**, 1479 (1982).
[6]L. Stenflo, Phys. Scr. **T50**, 15 (1994).
[7]R. P. Sharma, K. Ramamurthy, and M. Y. Yu, Phys. Fluids **27**, 399 (1984).
[8]H. Saleem and G. Murtaza, J. Plasma Phys. **36**, 295 (1986).
[9]V. K. Tripathi, C. S. Liu, and C. Grebogi, Phys. Fluids **22**, 301 (1979).
[10]C. S. Liu and V. K. Tripathi, Phys. Rep. **130**, 143 (1986).
[11]V. K. Jain and P. J. Christiansen, Phys. Lett. A **82**, 127 (1981).
[12]G. Praburam and A. K. Sharma, J. Plasma Phys. **48**, 3 (1992).
[13]W. L. Kruer, *The Physics of Laser Plasma Interactions* (Addision-Wesley, Reading, MA, 1987).
[14]E. C. Whipple, T. G. Northdrop, and D. A. Mendis, J. Geophys. Res. **90**, 7405, https://doi.org/10.1029/JA090iA08p07405 (1985).
[15]M. R. Jana, A. Sen, and P. K. Kaw, Phys. Rev. E **48**, 3930 (1993).
[16]V. W. Chow and M. Rosenberg, Planet. Space Sci. **43**, 613 (1995).
[17]V. W. Chow and M. Rosenberg, Planet. Space Sci. **44**, 465 (1996).
[18]A. Barkan, N. D'Angelo, and R. L. Merlino, Planet. Space Sci. **43**, 905 (1995).
[19]N. D'Angelo, Planet. Space Sci. **38**, 1143 (1990).
[20]R. L. Merlino, A. Barkan, C. Thompson, and N. D' Angelo, Phys. Plasmas **5**, 1607 (1998).
[21]B. Song, D. Suszcynsky, N. D'Angelo, and R. L. Merlino, Phys. Fluids **B1**, 2316 (1989).
[22]S. C. Sharma and M. P. Srivastava, Phys. Plasmas **8**, 679 (2001).
[23]S. C. Sharma, K. Sharma, and A. Gahlot, Phys. Plasmas **20**, 053704 (2013).
[24]S. C. Sharma and A. Gahlot, Phys. Plasmas **17**, 023708 (2010).
[25]V. W. Chow and M. Rosenberg, Phys. Plasmas **3**, 1202 (1996).
[26]D. M. Suszcynsky, N. D'Angelo, and R. L. Merlino, J. Geophys. Res. **94**, 8966, https://doi.org/10.1029/JA094iA07p08966 (1989).
[27]J.-X. Ma and M. Y. Yu, Phys. Plasmas **1**, 3520 (1994); Phys. Rev. E **50**, R2431 (1994).
[28]S. V. Vladimirov, K. N. Ostrikov, and M. Y. Yu, Phys. Rev. E **60**, 3257 (1999).
[29]K. N. Ostrikov, S. Kumar, and H. Sugai, Phys. Plasmas **8**, 3490 (2001).
[30]S. V. Vladimirov, K. Ostrikov, M. Y. Yu, and G. E. Morfill, Phys. Rev. E **67**, 036406 (2003).
[31]M. Rosenberg, E. Thomas, Jr., and R. L. Merlino, Phys. Plasmas **15**, 073701 (2008).
[32]M. S. Barnes, J. K. Keller, J. C. Forster, J. A. O'Neill, and D. K. Coultas, Phys. Rev. Lett. **68**, 313 (1992).
[33]A. Gahlot, R. Walia, J. Sharma, S. C. Sharma, and R. Sharma, Phys. Plasmas **20**, 013706 (2013).
[34]K. N. Ostrikov, M. Y. Yu, S. V. Vladimirov, and O. Ishihara, Phys. Plasmas **6**, 737 (1999).
[35]K. Ostrikov, S. V. Vladimirov, M. Y. Yu, and G. E. Morfill, Phys. Plasmas **7**, 461 (2000).
[36]A. Gahlot, Prog. Electromagn. Res. **57**, 63 (2017).
[37]R. Annou and V. K. Tripathi, Phys. Plasmas **5**, 60 (1998).
[38]J. X. Ma and P. K. Shukla, Phys. Plasmas **2**, 1506 (1995).
[39]P. K. Shukla and S. V. Vladimirov, Phys. Plasmas **2**, 3179 (1995).
[40]S. V. Vladimirov, Phys. Plasmas **1**, 2762 (1994).
[41]R. Annou and V. K. Tripathi, Phys. Lett. A **226**, 193 (1997).
[42]X. Wang and A. Bhattacharjee, Phys. Plasmas **3**, 1189 (1996).
[43]F. F. Chen, *Introduction to Plasma Physics* (Academic Press, New York, 1965).
[44]N. A. Krall and A. W. Trivelpiece, *Principles of Plasma Physics* (McGraw-Hill, New York, 1973).

Published under license by AIP Publishing

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Unpaved test sections reinforced with geotextile and geogrid

Meenakshi Singh [a,*], Ashutosh Trivedi [a], Sanjay Kumar Shukla [b]

[a] Department of Civil Engineering, Delhi Technological University, Delhi, India
[b] School of Building and Civil Engineering, Fiji National University, Fiji

ARTICLE INFO

Article history:
Received 10 January 2020
Received in revised form 3 March 2020
Accepted 11 March 2020
Available online xxxx

Keywords:
Geotextile
Geogrid
Unpaved road
Dynamic cone penetrometer

ABSTRACT

Geotextile and geogrid are used routinely to reinforce the unpaved roads to enhance the serviceability and stability of these roads. Field performance studies were conducted on unpaved test sections reinforced with geotextile and geogrid to demonstrate how geosynthetics improve the test section performance using dynamic cone penetrometer (DCP). The influence of geosynthetic reinforcement was investigated by placing the reinforcement layer at subgrade-base interface, thus helps to achieve more sustainable unpaved test section. Direct shear tests were conducted on soil samples reinforced with geotextile and geogrid to investigate the improvement in shear strength of geosynthetic reinforced soil. The results of DCP test were demonstrated as dynamic cone penetration index value (DCPI, mm/blow). DCPI values were observed in range from 1 to 7 mm/blow, corresponding to 500 mm depth of penetration for unreinforced and geogrid-reinforced test section and 312 mm depth of penetration for geotextile-reinforced test section respectively. A decrease in DCPI value was observed for the geosynthetic-reinforced unpaved test section as compared to the unreinforced unpaved test section. Direct shear test results shows a significant improvement in the shear strength of soil samples reinforced with geotextile.
© 2020 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the 3rd International Conference on Innovative Technologies for Clean and Sustainable Development.

1. Introduction

Geosynthetics have been used successfully to reinforce unpaved roads built on soft subgrade for many years. One of the common problems in road construction is dealing with weak subgrade. Unpaved road constructed on weak subgrade, generally causes severe damages to the pavement and requires regular maintenance and repair when they are subjected to the traffic load. In such situations geosynthetics can be used as a stabilization method of improvement of the serviceability of these roads by reinforcing them with geosynthetics. Geosynthetics placed either at the subgrade-base interface or within the granular layer may perform the desired function of separation, confinement or reinforcement. Geosynthetics, when used to serve as separation function then it will prevent the intermixing of granular material into the subgrade soil under cyclic loading. The primary function allocated to geotextiles is separation, but the reinforcement function may be acquired from both geogrids and geotextiles [1]. Several studies have shown the benefits of geotextile and geogrid reinforcement to improve the performance of unpaved roads [2,3].

The Dynamic Cone Penetrometer (DCP) equipment has been widely used for the pavement evaluation and in-situ strength of subgrade and pavement layers, where it has a strong relationship with various pavement design parameters such as the California Bearing Ratio (*CBR*) [4–6]. Various applications of DCP in pavement engineering is to measure the layer thickness of existing pavements, resilient modulus, condition of pavement, in-situ strength, layer stiffness and variation in base course and subgrade soil of existing pavement [7–9]. *CBR* tests performed by the conventional method are time-consuming and require costly equipment. To overcome this disadvantage, it is necessary to have an accurate, economical and reliable method for in-situ determination of *CBR* of subgrade soil and pavements. Geosynthetics can offer sustainable benefits over the traditional materials by permitting the use of lower quality aggregates, quick and easy to install, cheap to purchase, and easy to transport. Present study focuses on the use of DCP devices to examine the benefits of geotextile and geogrid reinforcement in unpaved test sections.

The soil-geosynthetic interaction plays an important role in many geotechnical engineering applications, especially in design

* Corresponding author.
E-mail address: meenakshisingh_2k14@dtu.ac.in (M. Singh).

https://doi.org/10.1016/j.matpr.2020.03.260
2214-7853/© 2020 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the 3rd International Conference on Innovative Technologies for Clean and Sustainable Development.

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260

and performance of soil structures reinforced with geosynthetics. This interaction between soil and geosynthetic depends on the properties of soil and geosynthetic reinforcement as well as the interrelationship between these materials. Numerical and experimental analyses have been reported in the literature which helps to generate the ability to understand the mechanism behind soil-geosynthetic reinforcement interaction [18].

2. Reinforcement mechanism

Geosynthetics can improve the performance of unpaved roads through the reinforcing mechanisms which include separation between base and subgrade, tensioned membrane effect, vertical restraint of the subgrade soil and lateral restraint of the base course [10]. Among these mechanisms other than separation between base and subgrade, lateral restraint of the granular material and vertical restraint of the subgrade soil becomes the dominant mechanisms responsible for enhancing the performance of unpaved roads within tolerable deformation. When the large deformation occurs in the subgrade then the tensioned membrane effect comes to play an important role. The mechanism behind the reinforcement function serve by geotextile is the tensioned membrane effect. If the subgrade undergoes large deformation due to the influence of traffic load, due to which the geosynthetic will exhibits a concave shape under the wheels. The deformed shape of geosynthetic is under tension and the vertical component of this tension developed in the geosynthetics will distribute the traffic load to a wider area in the subgrade. Geogrid improve the performance of base course through two closely related reinforcement mechanisms: improvement of wheel load distribution and lateral restraint of base course materials. Both the mechanisms were dependent of the interaction between the geosynthetic and granular material, which prevents lateral spreading of the granular material and imparts tensile stiffness to the base course. The role of granular layer in unpaved roads is to distribute the traffic loads. If the stiffness and thickness of the granular layer is sufficient, the maximum vertical stresses in the subgrade are below the stress level that causes excessive deformation in the subgrade soil. Geosynthetics at the interface of subgrade-base course is the most efficient and convenient location to get the maximum benefits in roadway construction [11,12]. The reinforcement mechanism responsible for road improvement related to the subgrade-base interaction is intermixing of base and subgrade material. Geotextiles are typically used to perform the separation function which prevents the intrusion of fine particles from the subgrade into the granular base course. Geogrids can also be used to perform the separation function at the subgrade-base interface, which provide some degree of separation depending on its aperture size. The interlocking action between the geogrid and granular material will be reduced if the fine particles from the subgrade soil intrude into the base course. Hence, the performance of unpaved roads will be affected because interlocking is the main mechanism of improvement of a pavement by a geogrid.

3. Materials

3.1. Subgrade soil

Locally available soil was used as the subgrade soil for field testing program. Representative samples of soil were excavated from the campus of DTU, Delhi. According to the Indian Soil Classification System, subgrade soil was classified as silty sand (SM). Standard Proctor compaction tests were conducted on subgrade soil to determine its maximum dry unit weight and optimum moisture content as presented in Table 1. The physical properties of the soil were determined in the laboratory and represented in Table 1.

Table 1
Properties of subgrade soil.

S. No.	Property	Value
1	Specific gravity	2.48
2	Plastic limit (%)	20
3	Liquid limit (%)	29
4	Plasticity index (%)	9
5	Optimum moisture content (%)	13.4
6	Maximum dry unit weight (Kn/m^3)	18.2

3.2. Granular material

The aggregates used in granular base course are designated as well-graded gravel (GW), according to the Indian Soil Classification System. Granular materials used in the study were obtained from the nearby quarry site. The size of the aggregate varies from 4.75 mm to 12.5 mm. The base course material has a specific gravity of 2.6. The granular material belongs to the grade-III of the base course material as per MORTH (Ministry of Road Transport and highways), New Delhi, India. The maximum dry unit weight obtained was 20.3 kN/m^3 at a water content of 5.2%.

3.3. Geosynthetic reinforcement

Geotextile and geogrid were used to reinforce the unpaved test sections and reinforced soil samples for conducting direct shear tests in this study as shown in Fig. 1. The specifications of geotextile and geogrid are presented in Table 2. Both geotextile and geogrid were introduced at the interface to understand the benefits of geosynthetics by conducting the field tests.

4. Field testing program

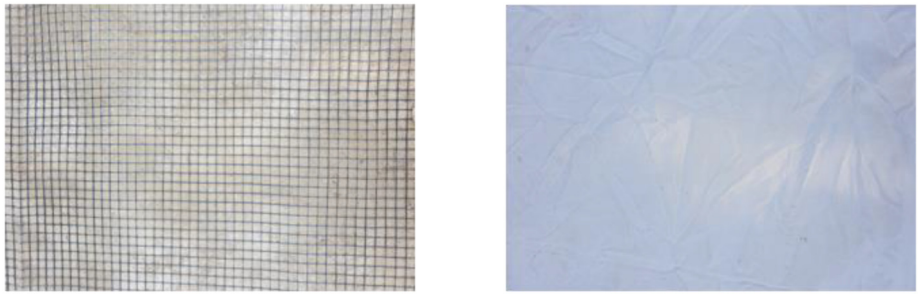
4.1. Test section construction

A total of three unpaved test sections were constructed over the natural soil ground located at the Delhi Technological University. Length and width of each unpaved test section is 400 mm. Fig. 2 shows the cross sections of the three unpaved test sections which consist of a 120-mm thick base course and 280-mm thick prepared soil subgrade with geotextile and geogrid reinforcement layer lay down at the interface of base-subgrade. Test section 1 was control section without geosynthetic reinforcement. Test section 2 and test section 3 were reinforced with geotextile and geogrid by placing them at the subgrade-base interface. The subgrade was prepared by compacting soil in two equal lifts at the optimum moisture content. After preparing the subgrade, geotextile and geogrid were placed over it for constructing geosynthetic-reinforced unpaved test sections. Geotextile and geogrid were used as geosynthetic reinforcement to prepare reinforced test section. Granular layer were laid over the geosynthetic reinforcement layer and compacted. Over the finished top of the granular layer surface dressing was laid. Field tests were performed at the centre of each unpaved test section using a dynamic cone penetrometer.

4.2. Dynamic cone penetrometer

The dynamic cone penetrometer (DCP) tests were performed according to the specifications of ASTM D6951-03. The DCP equipment consists of a 60° cone with a 20-mm base diameter connected at the bottom of a 16 mm diameter rod as shown in

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260



(a) (b)

Fig. 1. Geosynthetics used in the study (a) Geogrid and (b) Geotextile.

Table 2
Properties of geosynthetics used in the study.

S. No.	Property	Geotextile	Geogrid
1	Polymer type	Polypropylene	Polypropylene
2	Aperture size (mm)	0.075	30 × 30 (MD × CMD)
3	Tensile strength (kN/m)	45 × 34 (MD × CMD)	–
4	Stiffness at 0.5% strain (kN/m)	–	550 × 350 (MD × CMD)
5	Color	White	Black

Notes: MD: Machine direction; CMD: Cross machine direction

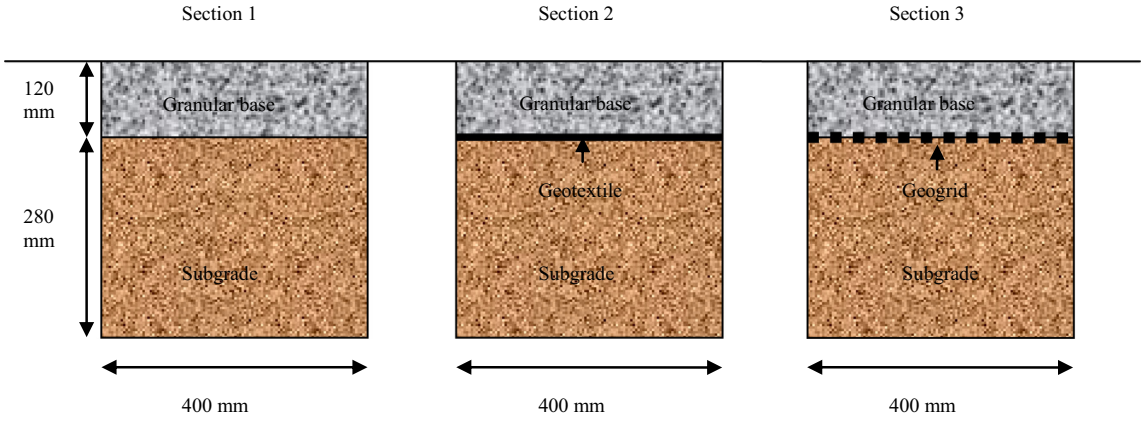


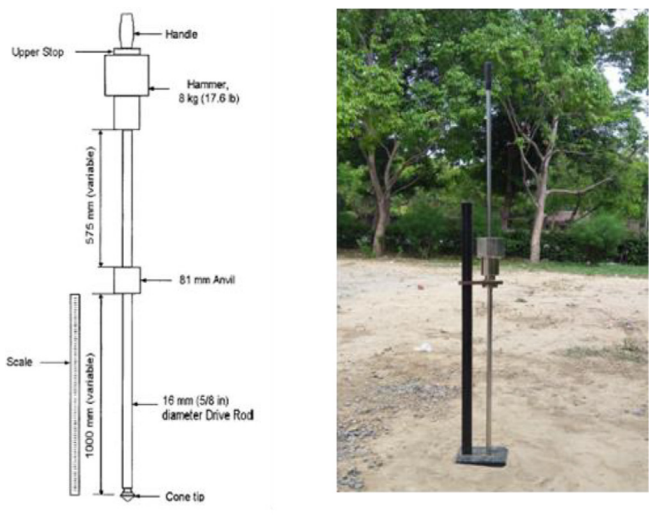
Fig. 2. Cross section of the unreinforced test section and geosynthetic reinforced test section.

Fig. 3. While conducting the DCP test the rod of equipment was held vertical. The hammer of weight 8 kg was dropped from a height of 575 mm and allowed to fall freely over the surface of unpaved test section to drive the cone into the test section layers by an impact load and measure the penetration depth for each number of blows. It is necessary to ensure that while raising the hammer upward it was not allowed to lift the cone upward before dropping the hammer. The penetration depth of the DCP device was recorded for each blow and these measurements are demonstrated as dynamic cone penetration index (DCPI). DCPI is described as the depth of penetration per blow of the hammer (mm/blow). The gradient of the graph plotted between penetration depth and cumulative number of blows represented as the DCPI [13]. DCP tests were conducted to measure the in-situ strength of granular base and subgrade material in terms of DCPI in mm/blow. The number of blows and the corresponding penetration depths were measured up to a penetration depth of 500-mm for unreinforced test section and geogrid reinforced test section. For geotextile reinforced test section, the number of blows was recorded up to a depth of penetration of 312 mm.

5. Direct shear test

Conventional direct shear test apparatus was used to investigate the influence of geosynthetic reinforcement on the reinforced soil specimen composite as shown in Fig. 4. The shear box consists of two boxes, one is upper box and another is lower box. Both the boxes are occupied by the same soil mass and upper box is fixed in the horizontal direction while, lower box allows the movement in the horizontal direction. Force is applied on the soil mass to slide along a horizontal failure plane under a constant rate of displacement. The load is applied normal to the plane of relative movement. Direct shear test apparatus consisted of load cells of capacity equal to 45 kN, and two linear variable differential transducers (LVDT). Vertical and horizontal displacements in the soil mass can be measured through LVDTs during the shearing process.

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260



(a) (b)

Fig. 3. Dynamic cone penetrometer (schematic image Source [19]).



Fig. 4. Direct shear test apparatus.

The equipment has Data Acquisition System where the measurements can be stored. A geosynthetic reinforcement layer was laid at the interface of soil mass between the two boxes and was tightly fixed to the lower box as shown in Fig. 5. The rate of shearing was kept at 1 mm/min in accordance with ASTM D 5321 [28].

6. Results and discussion

6.1. DCP test

Dynamic cone penetrometer test was performed at the centre location of both the geosynthetic reinforced test section and unreinforced test section. The DCP test results were used in distinguishing the granular base and subgrade and in determining the thickness of granular base and subgrade layer. DCPI value is useful to find the slope of the linear trend line, which mark the boundaries between the subgrade and granular layer due to the presence of geosynthetic reinforcement. Fig. 6 shows the DCP blow counts against the penetration depth for the unreinforced test section and geosynthetic-reinforced test sections. The graph plotted between the numbers of blows versus depth helps to observe the change in slope for each layer. It is difficult to find the exact position of the interface because a transition zone exists between the granular base and subgrade layer. It was observed that for the same numbers of blows the cone penetrates up to a depth of 312 mm for geotextile reinforced test section and 500 mm for geogrid reinforced test section. The DCPI values range from 2 mm/blow to 8 mm/blow for the granular layer and 1 mm/blow to 7 mm/blow for the subgrade layer. Overall DCPI values obtained for geotextile reinforced test section, geogrid reinforced test section, and unreinforced test sections are 1.91 mm/blow, 3.22 mm/blow and 6.29 mm/blow, respectively. A decrease in DCPI value was observed for both the geosynthetic-reinforced unpaved test sections as compared to unreinforced unpaved test section. Decreased values of DCPI indicate that the resistance to penetration is more in the geosynthetic-reinforced unpaved test section as compared with unreinforced unpaved test section. Hence, inclusion of geosynthetic reinforcement layer at the subgrade-base interface improves the performance of unpaved test section. Geogrid has higher strength and stiffness than geotextile instead of that, geotextile reinforced test section performed better than the geogrid reinforced test section. The dominating factor responsible for the improved performance of test section is reinforcing mechanism rather than the properties of geosynthetics [14].

6.2. DCPI–CBR correlation

The most common relationship exists in the analysis of DCP test data is between the DCPI value and *CBR* and existing correlation between them are summarized in Table 3. The in-situ strength of subgrade layer and granular layer of unreinforced and geosynthetic reinforced test section was estimated based on the field DCP data by the empirical correlation recommended by Webster et al. [15]. Conducting in-situ CBR tests in the pavement layers could be difficult, laborious and time consuming. Thus, these tests are rarely attempted and laboratory tests are preferred to determine the



(a) (b)

Fig. 5. Preparation of reinforced soil specimen composite (a) geogrid reinforcement (b) geotextile reinforcement.

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260

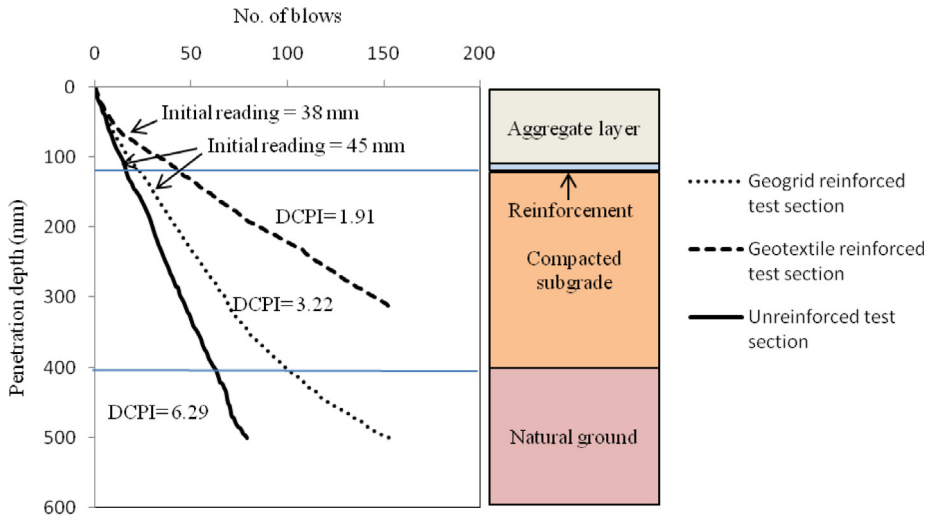


Fig. 6. DCP profile for unreinforced and geosynthetic-reinforced unpaved test sections.

Table 3
Relationship between DCP and CBR.

Standard Equation	a	b	Tested soil material	Description	Reference
$Log_{10}(CBR) = a - b \times log_{10}(DCPI)$	2.81	1.32	Granular and cohesive material	CBR and DCP experimental testing in the laboratory on samples prepared in standard CBR moulds	Harison, 1987 [20]
	2.53	1.14	Residual soil		Coonse, 1999 [22]
	2.70	1.35	Silty sand		Refeai and Suhaibani, 1997 [26]
	2.46	1.12	Granular and cohesive material	CBR experimental testing in the laboratory and DCP testing in the field	Livneh et al. 1994 [23]
	1.40	0.55	Aggregate base course material		Gabr et al. 2000 [21]
	2.44	1.07	Aggregate base course material		Ese et al. 1994 [24]
	2.62	1.27	Natural compacted soil		Kleyn, 1975 [19]
	2.56	1.15	Clayey subgrade	CBR experimental testing in the laboratory and field and DCP testing in the field	Smith and Pratt, 1983 [25]
	2.46	1.12	Unsurfaced soils and aggregate surfaced road	DCP testing in the field to measure road pavement strength	Webster et al. 1992 [15]
	2.48	1.057	Granular and cohesive material	Laboratory and field tests	TRL eqn [27]

Note: a = Constant; b = Coefficient

strength characteristics of pavement. DCP tests are advantageous for determining the layer thickness and in-situ strength of existing pavements. The repeatability and reproducibility of conducting CBR tests are typically poor as compared with the DCP tests [16]. Field *CBR* values have been obtained from the field DCPI values as per the correlation given by Webster et al. [15] and these values are presented in Table 4. Small DCPI value reflects a stronger soil with high *CBR* value whereas larger DCPI value suggests weaker soil of low *CBR* value. Increased *CBR* value was obtained for the geosynthetic-reinforced unpaved test section due to the benefits served by geosynthetic reinforcement [17].

6.3. Failure envelope

The shear stress in the soil sample reinforced with geosynthetic can be calculated by Eq. (1)

$$s = c' + \sigma' \tan\varphi \quad (1)$$

where, s is the shear stress of soil reinforced with geosynthetic, c' is the cohesion, σ' is the normal stress and φ is the friction angle of soil reinforced with geosynthetic.

Direct shear tests was conducted at normal stresses of 50 kN/m^2, 100 kN/m^2 and 150 kN/m^2 and shear stress at failure was recorded for three cases i.e unreinforced, geotextile reinforced soil sample and geogrid reinforced soil sample. Shear stresses obtained for geogrid and geotextile reinforced soil samples are 1.34 kN/m^2, 1.81 kN/m^2, 2.82 kN/m^2 and 1.52 kN/m^2, 2.10 kN/m^2, 2.68 kN/m^2 corresponding to normal stresses of 50 kN/m^2, 100 kN/m^2 and 150 kN/m^2. For unreinforced soil sample the shear stresses obtained are 1.22 kN/m^2, 1.51 kN/m^2, 2.36 kN/m^2 corresponding to normal stresses of 50 kN/m^2, 100 kN/m^2 and 150 kN/m^2. Failure envelopes were calculated by the linear trend of the data for all samples. Fig. 7 shows the plot of normal stresses against shear stresses with linear failure envelopes. Cohesion occurs in the reinforced soil and due to the inclusion of geosynthetic reinforcement in the soil sample there is a variation in failure envelopes when compared to each other. Good results for shear stress were obtained for the samples reinforced with geosynthetic as compared to the unreinforced soil sample. The improvement in the shear strength of soil sample through the inclusion of geosynthetic reinforcement can be calculated by the friction angle of soil-geosynthetic interface. Soil-geosynthetic interface frictional angle obtained for unreinforced soil sample, geogrid reinforced soil

Table 4
Summary of field *CBR* values of unreinforced and geosynthetic reinforced test section from DCP test.

Test section	Predicted in-situ *CBR* values from field DCPI values		
	Overall test section	Subgrade layer	Granular layer
Unreinforced test section	37.2	36.9	32.4
Geotextile reinforced test section	141.3	153.9	112.6
Geogrid reinforced test section	78.7	60.9	47.5

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260

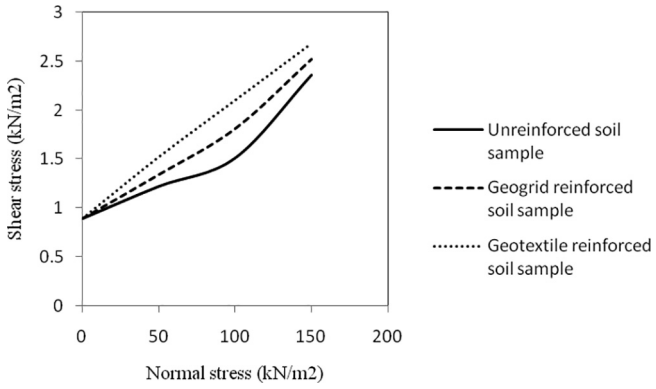


Fig. 7. Comparison of the failure envelopes for reinforced and unreinforced soil samples.

sample and geotextile reinforced soil sample are 11°, 21°, and 24°respectively.

7. Conclusions

In this study, dynamic cone penetrometer (DCP) tests were conducted on unreinforced unpaved test section and geosynthetic-reinforced unpaved test sections reinforced with geotextile and geogrid reinforcement layer at the interface of granular layer and subgrade. The DCP test results were used to estimate the influence of geosynthetic reinforcement on the performance of unpaved test section. Direct shear tests were conducted on unreinforced and geosynthetic reinforced soil sample to investigate the influence of geosynthetic on soil-geosynthetic interaction. Based on the field and direct shear testing, the following conclusions are drawn:

- DCPI value decreases for the reinforced test section as compared to the unreinforced test section due to the inclusion of geotextile and geogrid. The lowest DCPI value is obtained for the geotextile-reinforced test section.
- Greater penetration resistance is observed for geosynthetic reinforced test section as compared to the unreinforced test section.
- Geosynthetic reinforcement mechanism is the dominating factor responsible for the improved performance of unpaved test section, rather than the properties of geosynthetics.
- Greater shear stress improvement can be seen in the soil sample reinforced with geotextile by including a geosynthetic reinforcement layer in the soil mass.
- The friction angle (φ) obtained at the geosynthetic–soil interface lies between 20° to 25° for geosynthetic reinforced soil sample which is greater than the unreinforced soil sample.

CRediT authorship contribution statement

Meenakshi Singh: Conceptualization, Methodology, Investigation. **Ashutosh Trivedi:** Supervision, Validation. **Sanjay Kumar Shukla:** Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] E.V. Cuelho, S.W. Perkins, Geosynthetic subgrade stabilization – field testing and design method calibration, Transp. Geotech. 10 (2017) 22–34, https://doi.org/10.1016/j.trgeo.2016.10.002.
[2] M.Y. Abu-Farsakh, I. Akond, Q. Chen, Evaluating the performance of geosynthetic-reinforced unpaved roads using plate load tests, Int. J. Pavement Eng. 17 (2016) 901–912, https://doi.org/10.1080/10298436.2015.1031131.
[3] I.A.G. Góngora, E.M. Palmeira, Influence of fill and geogrid characteristics on the performance of unpaved roads on weak subgrades, Geosynth. Int. 19 (2012) 191–199, https://doi.org/10.1680/gein.2012.19.2.191.
[4] M. Boutet, G. Dore, J. Bilodeau, P. Pierre, Development of models for the interpretation of the dynamic cone penetrometer data, Int. J. Pavement Eng. 12 (3) (2011) 201–214, https://doi.org/10.1080/10298436.2010.488727.
[5] V. George, N.C. Rao, R. Shivashankar, PFWD, DCP and CBR correlations for evaluation of lateritic subgrades, Int. J. Pavement Eng. 10 (3) (2009) 189–199, https://doi.org/10.1080/10298430802342765.
[6] B.T. Nguyen, A. Mohajerani, Determination of CBR for fine-grained soils using a dynamic lightweight cone penetrometer, Int. J. Pavement Eng. 16 (2) (2015) 180–189, https://doi.org/10.1080/10298436.2014.937807.
[7] A.A. Amadi, S. Sadiku, M. Abdullahi, H.A. Danyaya, Case study of construction quality control monitoring and strength evaluation of a lateritic pavement using the dynamic cone penetrometer, Int. J. Pavement Res. Technol. 11 (5) (2018) 530–539, https://doi.org/10.1016/j.ijprt.2018.07.001.
[8] C. Chao, D. Lin, H. Luo, Y. Wang, W. Lo, Non-destructive evaluation of a city roadway for pavement rehabilitation: a case study, Int. J. Pavement Res. Technol. (2017), https://doi.org/10.1016/j.ijprt.2017.12.002.
[9] J. Chen, M. Hossain, T.M. Latorella, Use of Falling Weight deflectometer and dynamic cone penetrometer in pavement evaluation, Transp. Res. Rec. 1655 (1) (1999) 145–151.
[10] S.K. Shukla, An Introduction to Geosynthetic Engineering, CRC Press, London, 2016.
[11] B.M. Das, E.C. Shin, Strip foundation on geogrid-reinforced clay: Behaviorunder cyclic loading, Geotext. Geomembr. 13 (10) (1998) 657–666.
[12] J. Han, J.K. Thakur, Sustainable roadway construction using recycled aggregates with geosynthetics, Sustain. Cities Soc. (2014), https://doi.org/10.1016/j.scs.2013.11.011.
[13] H. Chennarapu, T.K. Garala, R. Chennareddy, U. Balunaini, G.V.N. Reddy, Compaction quality control of earth fills using dynamic cone penetrometer, J. Constr. Eng. Manage. 144 (9) (2018) 1–10.
[14] J. Kwon, E. Tutumluer, Geogrid base reinforcement with aggregate interlock and modeling of associated stiffness enhancement in mechanistic pavement analysis, Transp. Res. Rec. 2116 (2009) 85–95.
[15] Webster SL, Grau RH, Williams RP. Description and application of dual mass dynamic cone penetrometer. Vicksburg, MS, U.S. Army Engineer Waterways Experiment Station 1992. Report Number GL-92-3.
[16] J. Rolt, M.I. Pinard, Designing low-volume roads using the dynamic cone penetrometer, Transportation 169 (2016) 163–172.
[17] M. Singh, A. Trivedi, S.K. Shukla, Strength enhancement of the subgrade soil of unpaved road with geosynthetic reinforcement layers, Transp. Geotech. 19 (2019) 54–60, https://doi.org/10.1016/j.trgeo.2019.01.007.
[18] E.M. Palmeira, Soil–geosynthetic interaction: modelling and analysis, Geotext. Geomembr. (2009) 368–390, https://doi.org/10.1016/j.geotexmem.2009.03.003.
[19] Kleyn, E.G. (1975), the Use of the Dynamic Cone Penetrometer (DCP), Transvaal Roads Department, Report No. L2/74, Pretoria.
[20] A. Harison, Correlation between california bearing ratio and dynamic cone penetrometer strength measurement of soils, Proc. Inst. Civ. Eng. (1987) 832–844.
[21] Gabr, M. A., Hopkins, K., Coonse, J. and Hearne, T., (2000), DCP Criteria for Performance Evaluation of Pavement Layers, Journal of Performance of Constructed Facilities, Nov.2000, pp141-148.
[22] Coonse, J. (1999), Estimating California Bearing Ratio of COHESIVE piedmont Residual Soil using the Scala Dynamic Cone Penetrometor, Master's thesis, North Carolina State University, Raleigh, N.C.
[23] M. Livneh, I. Ishai, N.A. Livneh, Effect of vertical confinement on dynamic cone penetrometer strength values in pavement and subgrade evaluations, Transp. Res. Rec. 1473 (1994) 1–8.
[24] Ese, Dag, Myre, Jostein, Nos, Per Magne, and Vaernes, Einar. (1994), the Use of Dynamic Cone Penetrometer (DCP) for road strengthening design in Norway, Proc., Int. Conf. on Bearing Capacity of Rd. and Airfield. Pp. 3-22.
[25] R.B. Smith, D.N. Pratt, A field study of in situ California bearing ratio and dynamic cone penetrometer testing for road subgrade investigations, Australian Road Res. 13 (1983) 285–294.
[26] T. Al-Refeai, A. Al-Suhaibani, Prediction of CRR using dynamic cone penetrometer, J. King Saud Univ. Eng. Sci. 9 (2) (1997) 191–204.
[27] TRL (2004) UK DCP Version 2.2. TRL Limited, Wokingham, UK.
[28] ASTM Standard Test Method for Determining the Shear Strength of Soil-Geosynthetic and Geosynthetic-Geosynthetic Interfaces by D5321, 2012. Direct Shear.

Please cite this article as: M. Singh, A. Trivedi and S. K. Shukla, Unpaved test sections reinforced with geotextile and geogrid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.03.260

Using Hybridized techniques for Prediction of Software Maintainability using Imbalanced data

Ruchika Malhotra
Discipline of Software Engineering, Department of Computer Science and Engineering
Delhi Technological University, Delhi, India
ruchikamalhotra2004@yahoo.com

Kusum Lata
Department of Computer Science and Engineering,
Delhi Technological University, Delhi, India
kusumlata@dtu.ac.in

***Abstract*— Maintainability is an essential dimension of software quality. Software Maintainability Prediction (SMP) is gaining the attention of researchers to develop maintainable software systems. Early prediction of software maintainability aid the software practitioners to focus on those software modules or classes that requires high maintainability effort in the maintenance phase. However, the imbalanced distribution of training data is a challenging and serious problem that is encountered while developing prediction models for software maintainability. This paper apply oversampling methods namely: Adaptive Synthetic Oversampling technique (AdaS), Borderline-Synthetic Minority Oversampling technique (BSMOTE), Synthetic Minority Oversampling technique (SMOTE), and Safe-Level Synthetic Minority Oversampling technique (SSMOTE) to treat the imbalanced data before learning the models for software maintainability. We also investigate the effectiveness of hybridized techniques for learning the prediction models using three popular Apache datasets. The outcome of the study supports the use of investigated oversampling methods with hybridized techniques to develop effective prediction models for software maintainability.**

Keywords—Software Maintainability, Imbalanced data, Hybridized techniques

I. INTRODUCTION

Software maintenance is an essential phase in the life cycle of software development. This is post-delivery phase that initiates when the software product is delivered to the customer. As per literature, 40-60% of cost of software development is expended on maintenance and with the increasing complexity of software systems, this cost is ever escalating [1]. If we are able to predict software maintainability in the earlier during the software development, this would not only help to control the cost of software maintenance but also help to allocate the constraint resources of a software development organization in an optimal manner. Numerous machine learning (ML) and search-based techniques have been applied by the researchers to build prediction models for software maintainability [2-5]. Few studies have also examined hybridized techniques in this direction [6-8]. But while developing the prediction models for software maintainability, the imbalanced data problem is still unresolved. When the distribution of datapoints in a training data for a predictive modeling task is uneven, we say that the data is imbalanced [8]. For instance, in our case, the training data suffered from the imbalanced distribution of low maintainability (LM) and high maintainability (HM) datapoint. Here the LM datapoint means that software module or class representing this datapoint will require more changes in the lines of source code during the maintenance while a high maintainability class would require less changes. The prediction of LM classes with good accuracy is the prime concern of the software practitioners because these classes demand more maintenance effort in the maintenance phase. If we able to predict such classes early in the development phases, we can adopt a defensive design and thorough testing for such classes leading to good quality and maintainable software. But when the dataset used for developing the prediction models for software maintainability is imbalanced (uneven distribution of LM and HM), the prediction models cannot be trained effectively to predict the LM classes accurately. The LM classes correspond to minority classes, whereas HM classes correspond to majority classes. The accurate prediction of LM classes is the prime concern of this paper. To cater our aim, we apply data oversampling methods to balance the training data and then we develop SMP models by ML and hybridized techniques (HB) and compare their predictive performance. To pursue our aim, we address the following research questions (RQ's):

1. Do the investigated oversampling techniques, AdaS, BSMOTE, SMOTE, and SSMOTE enable improved performance of SMP models?
2. What is the relative performance of SMP models built by ML and HB techniques?

To answer above RQ's, we use three application packages of Apache. To assess the performance of the developed prediction models we use *Gmean* and *Balance* performance metrics. This study is organized as follows: section II describes the related work. Section III and Section IV present research method and experiment setup respectively. Section V presents the results. Section VI briefs the limitations of the study. We conclude the paper in section VII.

II. RELATED WORK

The imbalanced data problem is encountered in many real-life applications [10-12]. In software quality predictive modeling, the imbalanced data problem is confronted while predicting the faulty and change-prone classes. This problem is dealt by researchers by employing various data-level and algorithm-

978-1-7281-2791-0/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

level methods [9]. Pelayo and Dick [13] used SMOTE method on imbalanced defect datasets to improve the performance of software fault prediction (SFP) models. Sun et al. [14] applied coding-based ensembles to develop improved SFP models. Wang et al. [15] compared ensembles, data-resampling and threshold moving models for forecasting defective classes from imbalanced datasets. Tan et al. [16] employed resampling and updatable classification techniques to predict change-prone classes from imbalanced datasets. Malhotra et al. [17] used three data resampling methods for classifying the change-prone classes. Therefore, we see that in software quality predictive modeling, the imbalanced data is primarily addressed for SFP and predicting change-prone classes. However, based of the author's knowledge, the imbalanced data problem is untouched for SMP.

III. RESEARCH METHOD

This section briefly demonstrates the data collection process, dependent and independent variables used in the study, and the performance metrics used to assess the predictive performance of SMP models.

A. The Data Collection Process

In this study, we have used three application packages of Apache software, namely: Apache Bcel, Apache Lang, and Apache Jcs. For data collection corresponding to these applications, we have used Data Collection and Reporting Tool (DCRS), developed inhouse at Delhi Technological University [18]. The two successive versions of the application package are given input to the DCRS tool. The output obtained from DCRS was datapoints corresponding to the common class between the two analyzed versions. Each datapoint consist of value of eighteen object-oriented (OO) software metrics that we describe in the next section. In addition to OO metrics, each datapoint also includes lines of source code added, deleted, modified, and total lines of source code changed (*total_change*) from old version to new version of the software that was analyzed by DCRS tool. The datapoints obtained are then preprocessed to remove the outliers. We transformed the *total_change (*continuous variable) into *Maintainability* (binary variable with two values: LM and HM). The LM class requires more changes in the lines of source code as compared to HM class. Table 1 describes the specifics of the datasets that we have used. The specifics of the datasets include the name of the software, version analyzed, common classes between the analyzed versions (#CC), number of Low maintainability datapoints (#LM), number of High maintainability datapoints (#HM), and Imbalance ratio (IR). The IR is given as (#HM/#LM)*100. It is clearly evident from Table 1 that all three datasets used in the study are highly imbalanced.

Table I. Details of the software project.

Software	Version	# CC	#LM	#HM	IR
Apache Bcel	5.0 – 5.1	334	18	316	17.55
Apache Lang	2.4 – 2.6	220	27	193	7.14
Apache Jcs	1.2.6.5 – 1.2.7.9	197	27	170	6.29

B. Dependent and Independent Varaibles

The OO metrics are advocated as good predictors of software maintainability. Therefore, in this study we have used eighteen prevalent OO metrics describing different characteristics of a software system. The OO metrics used as independent in this study include Martin's metrics (Ca-Afferent coupling, Ce-Efferent coupling) [19], Henderson-Sellers metrics (CBM-Coupling Between Methods , LCOM3-Lack of cohesion in methods Henderson-Sellers version, AMC-Average Method Complexity) [20], QMOOD (Quality Metrics for Object oriented Design) metrics (NPM-Number of Public Methods for a class, MFA-Measure of Functional Abstraction, DAM-Data Access Metric, MOA-Measure of Aggregation, CAM-Cohesion Among Methods of Class) [21], and Chidamber & Kemerer metrics (WMC-Weighted methods per class, CBO-Coupling between object classes, LCOM-Lack of cohesion in methods, DIT-Depth of Inheritance Tree, NOC-Number of Children, RFC-Response for a Class) [22]. Few other OO metrics IC-Inheritance Coupling, cyclomatic complexity are also used in addition to above metrics. Due to space constraint, the definition of these metrics are not given in the study. The detailed description of these metrics can be referred from [23]. The dependent variable in this study is *Maintainability*. This is a binary variable that consists of two values: LM and HM as discussed in section III.

C. Performance Metrics

Accuracy is the commonly used performance metric to assess the predictive capability of prediction models. This metric is derived from the confusion matrix (Table II). However, this measure is biased towards *True Positive Rate* (*T.P.R*).

Table II. Confusion Matrix

		Actual Class
Predicted Class	TRUE Positives	FALSE Negatives
	FALSE Postives	TRUE Negatives

For the imbalanced domain, it is equally important to consider the *True Negative Rate (T.N.R)* as well. In this study, *Positive* correspond to HM class (majority class) instances, whereas Negative correspond to LM cass (minority class).

$$T.P.R = \frac{TRUE\ Postivies}{TRUE\ Positives + FALSE\ Negatives} \quad \text{(i)}$$

$$T.N.R = \frac{TRUE\ Negatives}{True\ Negatives + False\ Positives} \quad \text{(ii)}$$

We have used stable performance metric in this study, that consider T.P.R as well as T.N.R. These performance metrics are *Gmean* and *Balance. Gmean* is the geometric mean of *T.P.R* and *T.N.R.* The *Balance* performance metric measures

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

the Euclidian distance between the pair of *T.P.R* and *FALSE positive rate (F.P.R)* and an optimal value of *T.P.R* and *F.P.R*.

$$Gmean = \sqrt{T.P.R * T.N.R} \quad \text{(iii)}$$

$$Balance = 1 - \sqrt{\frac{(0 - F.P.R)^2 + (1 - T.P.R)^2}{2}} \quad \text{(iv)}$$

$$\text{where } F.P.R = \frac{FALSE\ positives}{FALSE\ positives + TRUE\ negatives}$$

IV. EXPERIMENTAL SETUP

This section describes the experimental setup for developing the prediction models for software maintainability. We have used three Apache datasets for experimentation. As shown in Table I, the datasets are highly imbalanced, therefore, before training the models we have employed oversampling techniques namely: AdaS, BSMOTE, SMOTE, and SSMOTE to balance the data distribution of instances of LM and HM classes.

For training the prediction models, we applied three ML techniques in commonly used literature and four HB techniques. The brief description of the oversampling and learning algorithms is given in the next subsection. The Ten-Fold Cross-validation (TFCV) strategy is used for training the models. Using TFCV, we divided the available training data into ten equal-sized partitions. Nine of the ten partitions were used for the training, whereas the model testing was carried out on the remaining partition. This process was repeated ten times so that each partition get equal chance to participate in train-test process.

A. Learning techiques

For developing prediction models for software maintainability, we have applied three ML techniques and four hybridized techniques. These techniques are briefly described as follows:

AdaBoost (AB): This is a boosting technique that learns a number of weak classifiers and improves their performance using weight updating [24].

Bagging (BG): In this technique, the training data is partitioned and a classifier is learned from each partition. The predictions of individual classifiers are then aggregated using majority voting to obtain the outcome [25].

C4.5: This is a decision tree classifier that makes use of gain-ratio for selection of features at the internal nodes [25].

Genetic Fuzzy Adaboost (GFS-AB), Genetic Fuzzy Logitboost (GFS-LB): These technique combines boosting techniques, namely Adaboost and Logitboost respectively with genetic algorithm (GA). GA is used to obtain best fuzzy rules that act as base classifiers to Adaboost and Logitboost. The sum of votes are accounted for the single decision. [26] [27].

Genetic Fuzzy MaxLogitboost (GFS-MaxLB): Unlike GFS-AB and GFS-LB, GFS-MaxLB is based on "single winner" strategy to output the final label [28]

Particle Swarm Optimization – Linear Discriminant Analysis (PSOLDA): This technique combines Particle Swarm Optimization (PSO) with Linear Discriminant Analysis (LDA) where PSO is used to specify the beneficial features from the training dataset that are then passed to LDA to build the classification model [29]. For developing the SMP models using these learning techniques, we have used Knowledge Extraction based on Evolutionary Learning (KEEL) tool [30].

B. Oversampling techiques

To balance the training data, the following oversampling techniques have been used in this study.

Adaptive Synthetic Oversampling technique (AdaS): This technique oversamples the minority class instances according to their difficulty level in learning. If a minority instance is difficult to learn, more samples are created corresponding to it and vice versa [31].

Synthetic Minority Oversampling technique (SMOTE): This technique creates synthetic samples of minority class whereas the number of majority class samples remain unchanged. Synthetic samples are generated along the line joining minority class instances with their *k*-nearer neighbors [32].

Borderline-Synthetic Minority Oversampling technique (BSMOTE): In this technique, only borderline minority class instances are in oversampled using SMOTE. A borderline minority class instance is a datapoint whose *k*-nearer neighbors belongs to majority class [33].

Safe-Level Synthetic Minority Oversampling technique (SSMOTE): This technique determines the safe-level of a minority class instance before oversampling. For a minority class instances, I, safe-level (I) = number of minority class instances in the *k*-nearer neighbors of I [34].

V RESULTS

A. Do the investigated oversampling techniques, AdaS, BSMOTE, SMOTE, and SSMOTE enable improved performance of SMP models?

To answer this research question, we develop SMP models, first using original imbalanced datasets with the application of ML and HB described in section IV. The performance of developed models on original imbalanced datasets in terms of performance metrics *Gmean* and *Balance* is reported in Table III. To develop the prediction models, we divide the dataset into train-test using TFCV. We then apply the oversampling techniques described in section IV and develop the models. Table III describes the performance of SMP models using original imbalanced data in terms of *Gmean* and *Balance*.

As shown in Table III, for imbalanced scenario, the predictive capability SMP models are not good in terms of *Gmean* and Balance. For Apache Bcel dataset *Gmean* of PSOLDA is 71.86 and *Balance* is 70.40 and for all other ML and HB, *Gmean* and *Balance* values are even less than 60. For Apache Lang dataset, in imbalanced scenario, BG technique has the best *Gmean* of 79.72 and *Balance* value of 76.20. For Apache Lang dataset, all learning techniques except AB, BG, and C4.5 reported *Gmean* less than 70. Also, for this dataset except BG

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

and C4.5, *Balance* values of all other learning techniques are less than 70. For imbalanced Apache Jcs dataset, *Gmean* and *Balance* values of all learning techniques are less than 70. GFS-LB technique has given best *Gmean* of 69.86 and *Balance* of 65.70 for Apache Jcs.

Table III. *Gmean* and Balance of SMP models on imbalanced datasets

Apache Bcel		
Learning Techniques	*Gmean*	*Balance*
GFS-AB	52.54	48.93
GFS-LB	40.70	41.07
GFS-MaxLB	46.99	45.00
PSOLDA	71.86	70.40
AB	57.18	52.84
BG	57.28	52.85
C4.5	33.17	37.14
Apache Lang		
GFS-AB	60.22	55.45
GFS-LB	65.80	60.67
GFS-MaxLB	68.12	63.25
PSOLDA	54.43	50.24
AB	70.12	65.76
BG	79.72	76.20
C4.5	76.84	73.44
Apache Jcs		
GFS-AB	62.88	58.05
GFS-LB	69.86	65.70
GFS-MaxLB	59.23	55.32
PSOLDA	62.88	58.05
AB	64.07	60.35
BG	65.68	60.66

Table IV. *Gmean* values of SMP models after oversampling

Apache Bcel				
Learning Techniques	AdaS	SMOTE	BSMOTE	SSMOTE
GFS-AB	75.29	76.64	73.11	76.49
GFS-LB	75.74	77.38	69.01	77.38
GFS-MaxLB	77.81	73.12	74.76	75.53
PSOLDA	76.93	80.90	76.21	78.67
AB	73.60	74.37	75.53	77.27
BG	73.29	75.61	75.27	76.72
C4.5	72.82	74.50	75.27	77.13
Apache Lang				
GFS-AB	79.34	82.28	82.53	82.66
GFS-LB	81.93	86.44	87.42	84.97
GFS-MaxLB	83.97	83.12	83.58	84.46
PSOLDA	83.67	85.60	82.53	87.96
AB	72.70	78.63	74.75	82.53
BG	75.04	81.01	79.07	83.26
C4.5	76.62	84.46	83.35	83.26
Apache Jcs				
GFS-AB	81.81	83.71	83.08	83.71
GFS-LB	81.33	81.33	78.94	79.85
GFS-MaxLB	85.24	81.62	79.21	83.46
PSOLDA	87.11	87.78	88.12	87.11
AB	80.03	79.06	75.28	78.66
BG	79.21	78.94	78.80	78.88
C4.5	80.89	80.03	70.48	78.88

Table IV and Table V, respectively show the *Gmean* and *Balance* results of SMP models after employing oversampling techniques. As depicted in Table IV and Table V, there positive increase in the values of *Gmean* and *Balance* after oversampling. All of the learning techniques on all three datasets, have given *Gmean* and *Balance* values greater than 70 (Table IV and Table V).

Table V. *Balance* values of SMP models after oversampling

Apache Bcel				
Learning Techniques	AdaS	SMOTE	BSMOTE	SSMOTE
GFS-AB	75.15	76.33	68.46	76.21
GFS-LB	75.56	76.94	64.49	76.94
GFS-MaxLB	75.55	71.12	71.85	72.10
PSOLDA	73.30	77.94	76.18	77.13
AB	73.57	71.69	72.10	75.30
BG	73.27	74.37	72.02	75.02
C4.5	72.81	71.75	72.02	75.23
Apache Lang				
GFS-AB	79.29	81.81	81.99	82.63
GFS-LB	81.93	85.65	86.19	84.62
GFS-MaxLB	83.81	81.06	81.23	83.23
PSOLDA	83.05	85.59	81.99	87.93
AB	71.45	75.88	70.91	81.99
BG	73.99	78.51	76.02	82.5
C4.5	75.99	83.23	81.15	82.5
Apache Jcs				
GFS-AB	80.84	83.09	83.02	83.09
GFS-LB	81.04	81.33	78.50	79.74
GFS-MaxLB	85.24	81.62	78.72	83.39
PSOLDA	82.95	83.78	84.19	82.95
AB	79.32	77.61	72.87	78.28
BG	78.72	78.50	77.46	78.86
C4.5	80.69	79.32	67.69	78.86

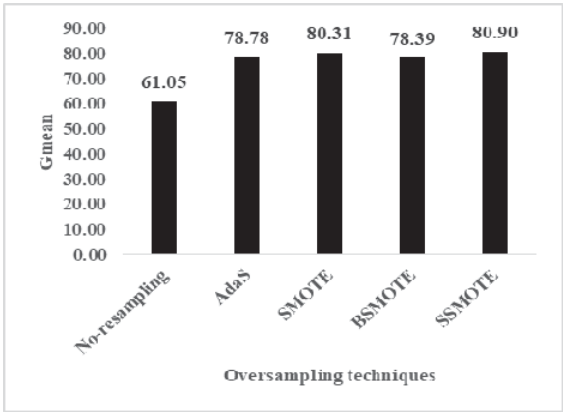


Fig. 1. Average *Gmean* values after oversampling

Average *Gmean* and *Balance* values of software maintainability predcition models after employing oversampling techniques along with no resampling situation are presented in Fig. 1 and Fig. 2. It is evident from Fig. 1 and 2 that there is vast improvement in the perfromance of predcition models after employing oversampling techniques. We, further validate our results with the help of Friedman test. The mean ranking obtained after oversampling techniques along with no resanpling situation in given in Table VI for both *Gmean* and *Balance.* The obtained p-values was 0.00 both for *Gmean* and *Balance*, which shows that the results of Friedman test are
significant. As shown in Table VI, the SSMOTE technique has obtained best rank for both *Gmean* and *Balance* and SMOTE has obtained secong best rank. The no-resampling

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

case has obtained worst rank. Therefore, we conclude that investigated oversampling techniques, AdaS, BSMOTE, SMOTE, and SSMOTE enable improved performance of SMP models.

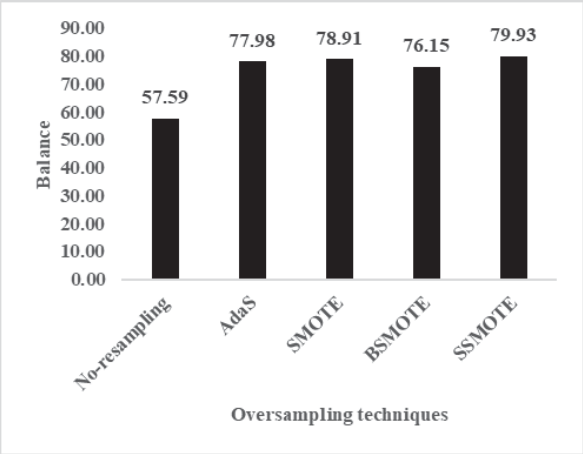


Fig. 2. Average *Gmean* values after oversampling

Table VI. Friedman Ranks after oversampling

Performance metric	Rank #1	Rank #2	Rank #3	Rank #4	Rank #5
Gmean	SSMOTE (4.07)	SMOTE (3.83)	AdaS (3.05)	BSMOTE (2.90)	No-resampling (1.14)
Balance	SSMOTE (4.21)	SMOTE (3.67)	AdaS (3.31)	BSMOTE (2.71)	No-resampling (1.10)

B. What is the relative performance of SMP models built by ML and HB techniques?

To answer this RQ, we conducted statistical analysis using Friedman test and compared the performance of ML and HB techniques with respect to cumulative *Gmean* and *Balance* of all three datasets used in the study. The results of Friedman ranking for *Gmean* and *Balance* are reported in Table VII and Table VIII respectively. The p-value obtained was 0.00 both for *Gmean* and *Balance* which means that the results of the Friedman test for performance of ML and HB techniques are significant. As shown in Table VII and Table VIII, the best rank is obtained by PSOLDA technique followed by GFS-LB, GFS-MaxLB, and GFS-AB in terms of both of the performance metrics. Further, it is to be noted that best four ranks are obtained by HB techniques with respected to *Gmean* as well as for *Balance*. The ML techniques have obtained lowest three ranks among the investigated techniques. Therefore, we conclude that HB techniques have better predictive capability than those of the ML techniques for predicting software maintainability.

Table VII. Friedman Ranks of ML and HB techniques using *Gmean*

Rank #1	Rank #2	Rank #3	Rank #4	Rank #5	Rank #6	Rank #7
PSOLDA (6.46) (HB)	GFS-LB (4.92) (HB)	GFS-MaxLB (5.48) (HB)	GFS-AB (4.04) (HB)	C4.5 (3.21) (ML)	BG (2.54) (ML)	AB (2.25) (ML)

Table VIII. Friedman Ranks of ML and HB techniques using *Balance*

Rank #1	Rank #2	Rank #3	Rank #4	Rank #5	Rank #6	Rank #7
PSOLDA (6.17) (HB)	GFS-LB (5.17) (HB)	GFS-MaxLB (4.50) (HB)	GFS-AB (4.50) (HB)	C4.5 (2.96) (ML)	BG (2.54) (ML)	AB (2.25) (ML)

VI THREATS TO VALIDITY

The SMP models are developed using TFCV and results are statistically validated with the help of Freidman test. So, there is no threat to conclusion validity. In this study, maintainability prediction models are developed by treating the imbalanced datasets, and use of stable performance metrics (*Gmean* and *Balance*) has reduced the threat to internal validity. This paper does not face any construct validity threat as the predictors used in the paper has widespread use in predictive modeling in software engineering. This study has used dataset pertaining to JAVA software system. There may exist an external validity threat that the results may differ for software written in programming languages other than JAVA programming language.

VI CONCLUSION AND FUTURE WORK

Early prediction of low maintainability classes is very much important to develop a good quality and cost-effective software. The imbalanced data poses difficulty for effective training of the prediction models. Therefore, in this direction the study effectively treated the imbalanced data by employing four oversampling techniques. The SMP models are developed using HB and ML techniques and the results are accessed with the help of two stable performance metrics, *Gmean* and *Balance*. The study advocate that the investigated oversampling techniques significantly improves the predictive capability of the SMP models. Particularity, SSMOTE is very effective in treating the imbalanced datasets. Further, HB techniques outperformed as compare to ML techniques in terms of *Gmean* and *Balance* for developing prediction models of software maintainability. The predictive performance of PSOLDA technique is outstanding in accordance to Friedman test.

In this study, we have used three datasets pertaining to Apache application packages. We will reproduce this study in future with more datasets from diverse domain in order to have more generalized and comprehensive results.

REFERENCES

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

[1] I. Sommerville, Software Engineering, Addison-Wesley, 1995.

[2] C. Van Koten and A. Gray, "An application of bayesian network for predicting object-oriented software maintainability," *Information and Software Technology*, vol. 48, no. 1, 2006, pp. 59-67.

[3] Y. Zhou and H. Leung, "Predicting object-oriented software maintainability using multivariate adaptive regression splines," *Journal of Systems and Software*, vol. 80, no. 8, 2007, pp. 1349-1361.

[4] K. K. Aggarwal, Y. Singh, A. Kaur and R. Malhotra, "Application of artificial neural network for predicting maintainability using object-oriented metrics." *Transactions on Engineering, Computing and Technology*, vol. 15, 2006, pp. 285-289.

[5] T. S. Quah and & M. M. T. Thwin, "Application of neural networks for software quality prediction using object-oriented metrics, " *In Proceeding of IEEE International Conference on Software Maintenance,* ICSM 2003, pp. 116-125.

[6] R. Malhotra and K. Lata, "An Exploratory Study for Predicting Maintenance Effort using Hybridized Techniques," *In Proceedings of 10th ACM Innovations in Software Engineering Conference*, Feburary 2017, pp. 26-33.

[7] L. Kumar, S. K. Rath, S. K. Rath, "Neuro–genetic approach for predicting maintainability using Chidamber and Kemerer software metrics suite," *In Recent Advances in Information and Communication Technology*, Springer, 2015, pp. 31-40.

[8] L. Kumar, S. Lal and L. B. Murthy, "Estimation of maintainability parameters for object-oriented software using hybrid neural network and class level metrics. *International Journal of System Assurance Engineering and Management*, 2018, pp.1–31.

[9] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications ", *Expert Systems with Applications*, vol. 73, 2017, pp. 220-239.

[10] A. C. Bahnsen, A. Stojanovic, D. Aouada and & B. Ottersten, "Cost sensitive credit card fraud detection using Bayes minimum risk," *In IEEE 12th International conference on machine learning and applications* , December 2013, pp. 333-338.

[11] A. Abbasi and H. Chen, "A comparison of fraud cues and classification methods for fake escrow website detection," *Information Technolog and Management*, vol. 10, 2009, pp. 83-101.

[12] L. Song, D. Li, X. Zeng, Y. Wu, L. Guo and & Q. Zou, "nDNA-prot: identification of DNA-binding proteins based on unbalanced classification", *BMC bioinformatics*, vo. 15 no. 1, 2014, 298.

[13] L. Pelayo, S. Dick, "Applying novel resampling strategies to software defect prediction" *In IEEE 2007 Annual Meeting of the North American Fuzzy Information Processing Society*, NAFIPS 2017, pp. 69-72.

[14] Z. Sun, Q. Song and X. Zhu, " Using coding-based ensemble learning to improve software defect prediction," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, 2012, pp. 1806-1817.

[15] S. Wang, X. Yao, X. "Using class imbalance learning for software defect prediction"' *IEEE Transactions on Reliability*, vol. 62, no. 2, 2013, pp. 434-443.

[16] M. Tan, L. Tan, S. Dara ans C. Mayeux, "Online defect prediction for imbalanced data," I*n 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering* ,vol. 2, May 2015, pp. 99-108.

[17] R. Malhotra and M. Khanna, "An empirical study for software change prediction using imbalanced data," *Empirical Software Engineering*, vol. 22, no. 6, 2017, pp. 2806-2851.

[18] R. Malhotra, N. Pritam, K. Nagpal and P. Upmanyu, (2014, September). "Defect collection and reporting system for git based open source software," *In IEEE International Conference on Data Mining and Intelligent Computing*, ICDMIC September 2014, pp. 1-7.

[19] R. C. Martin, "Object oriented design quality metrics," *Report on Object Analysis & Design*, 1995.

[20] B. Henderson-Sellers, "Object-oriented metrics: measures of complexity. 1995, Prentice-Hall Inc.

[21] J. Bansiya, C. G. Davis, "A hierarchical model for object-oriented design quality assessment," *IEEE Transactions on software engineering*, vol. 28, no. 1, 2002, pp. 4-17.

[22] S. R. Chidamber and C. F. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on software engineering*, vol. *20, no.* 6, 1994, pp. 476-493.

[23] http://gromit.iiar.pwr.wroc.pl/p_inf/ ckjm/metric.html

[24] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. *55*, no.1, 1997, pp. 119-139.

[25] J. R. Quinlan, " C4. 5: programs for machine learning," 2014, Elsevier.

[26] M. J. Del Jesus, F. Hoffmann, L.J. Navascués and L. Sánchez, "Induction of fuzzy-rule-based classifiers with evolutionary boosting algorithms," *IEEE Transactions on Fuzzy Systems*, vol. *12, no.* 3,2004, pp. 296-308.

[27] J. Otero, and L. Sánchez, "Induction of descriptive fuzzy classifiers with the Logitboost algorithm," *Soft Computing*, vol. 10, no. 9, 2006, pp. 825-835.

[28] L. Sánchez, and J. Otero, "Boosting fuzzy rules in classification problems under single□winner inference. *International Journal of Intelligent Systems*, vol. *22*, no. 9, 2007, pp. 1021-1034.

[29] S. W. Lin and S. C. Chen, "PSOLDA: A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis," *Applied Soft Computing*, vol. 9, no. 3, 2009, pp. 1008-1015.

[30] http://www.keel.es

[31] H. He, Y. Bai, E. A. Garcia and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," In *2008 IEEE International Joint Conference on Neural Networks,* June 2008, pp. 1322-1328.

[32] N. V. Chawla, K.W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique*," Journal of artificial intelligence research,* vol. 16, 2002, pp. 321-357.

[33] H. Han, W. Y. Wang and B. H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," In *International conference on intelligent computing*. Springer, Berlin, Heidelberg, August 2005, pp. 878-887.

[34] C. Bunkhumpornpat, K. Sinapiromsaran and C. Lursinsap, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," *In Pacific-Asia conference on knowledge discovery and data mining*, Springer, Berlin, Heidelberg, April 2009, pp. 475-482.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 08:50:51 UTC from IEEE Xplore. Restrictions apply.

2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)

VDTA Based Fractional Order Floating Inductor and its Applications

Megha Jain
Department of Electronics and Communication
Delhi Technological University
Delhi, India
megha98jn@gmail.com

Reshma S Kumar
Department of Electronics and Communication
Delhi Technological University
Delhi, India
gsreji1@gmail.com

Lakshmi Nair
Department of Electronics and Communication
Delhi Technological University
Delhi, India
lakshmi2798@gmail.com

Rajeshwari Pandey
Department of Electronics and Communication
Delhi Technological University
Delhi, India
rpandey@dce.ac.in

***Abstract*—In this paper, a fractional order floating inductor using two Voltage Differencing Transconductance Amplifiers (VDTAs) and a fractional order capacitor (FC) has been introduced. Further, three applications of Fractional Inductor (FI) namely fractional order low pass filter (FLPF), fractional order high pass filter (FHPF) and fractional order band pass filter (FBPF) have been presented. The proposed designs are verified through SPICE simulations using TSMC 180 nm CMOS model parameters and FC is realized using RC Domino ladder network truncated to 11-12 blocks.**

Keywords—VDTA, FC, FI, filter

I. INTRODUCTION

In recent times, the popularity of fractional calculus has risen owing to the fact that it serves as a powerful tool to model real world applications which are generally of fractional origin and not always of the integral order. This shift from the traditional integer order systems to fractional ones has thus facilitated enhanced flexibility, freedom and provided more room for optimization. The circuit theory relies on three basic passive circuit elements, i.e. Resistor, Capacitor and Inductor, hence many attempts have been made in the past few decades to realize their fractional order counterparts due to which the phase difference between the current and voltage is expected to show a linear dependence on the fractional order and provide a variation over full 2π phase [1].

This paper aims at presenting the floating fractional inductor employing VDTA and the already reported fractional order capacitor [2].

The choice of VDTA as the active circuit element is based on the advantages that it offers such as two different tunable values of transconductance and its usefulness in achieving compact structure realizations suitable for fully integrated circuit designs.

This approach of simulating passive inductor characteristics using active circuit elements is motivated by the fact that conventional coil type inductors suffer from resistive losses and capacitive coupling losses whereas their simulated active counterparts provide electronic tunability and controllability with relatively broad range of inductance values at a low cost and comparatively less space [3].

In open literature, several studies that simulate the fractional capacitors and inductors have been reported using various active elements like the Current Conveyors [4], Operational Amplifiers (Op-Amps) [5], operational transconductance amplifiers (OTAs) [6], Current feedback Operational Amplifiers (CFOAs) [7], Current Controlled Current Feedback Operational Amplifier (CCCFOAs) [3] as summarized in Table I which presents a comparison of the implemented FI with the existing FI designs in terms of the active and passive components utilised for its simulation. Consequently, this reveals the significance of the proposed configuration which employs only two VDTAs and an FC, emerging as the design with minimum number of circuit elements for the fractional inductor (FI) simulation.

Rest of the paper is structured as below: Section II highlights the design methodology of VDTA and FC based proposed FI. Section III presents the practicability of the devised FI in basic filter circuit configurations. Finally, Section IV compiles all the simulated plots and curves to draw constructive inferences and conclusions henceforth which have been presented in section V.

II. THE PROPOSED CONFIGURATION

The impedance and phase function of an FC [8] are expressed as (1) and (2) respectively

$$Z_C(s) = \frac{1}{s^{\alpha} C_{\alpha}} \tag{1}$$

TABLE I. COMPARISON OF PROPOSED FI WITH EXISTING FI DESIGN

Reference	Inductance Type	Active Block	No. of Active Devices	No. of Resistors	No. of Integer Order Capacitors	No. of FCs
[3]	Grounded	CCCFOA	2	0	0	1
[5]	Floating	Op-Amp	2	4	1	0
[6]	Floating	OTA	3	3	1	0
[7]	Grounded	CFOA	1	2	0	1
[11]	Grounded	CMOS-RC network	3	0	0	1
Proposed	Floating	VDTA	2	0	0	1

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-7281-5475-6/20/$31.00 ©2020 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.

$$\left|Z_C(j\omega)\right| = \frac{1}{\omega^{\alpha} C_{\alpha}} \tag{2}$$

The capacitance is represented in units of $\mho/s^{\alpha}$, where α is the order of the FC.

Similarly, the impedance and phase function of an FI of fractional order β are given by (3) and (4) respectively.

$$Z_L(s) = s^{\beta} L_{\beta} \tag{3}$$

$$\left|Z_L(j\omega)\right| = \omega^{\beta} L_{\beta} \tag{4}$$

The circuit symbol of VDTA [9] as shown in Fig. 1 has five terminals. While V_p and V_n represent the high impedance input terminals, the Z, X_+ and X_- are high impedance current output terminals. The terminal characteristics of VDTA are represented by matrix (5).

$$\begin{bmatrix} I_Z \\ I_{X+} \\ I_{X-} \end{bmatrix} = \begin{bmatrix} g_{m1} & -g_{m1} & 0 \\ 0 & 0 & g_{m2} \\ 0 & 0 & -g_{m2} \end{bmatrix} \begin{bmatrix} V_{V_p} \\ V_{V_n} \\ V_Z \end{bmatrix} \tag{5}$$

In (5) g_{m1} and g_{m2} represent the transconductance gains of the first and second transconductance stages of VDTA respectively. The g_{m1} and g_{m2} can be tuned electronically with the help of bias currents of the transistors.

The proposed FI of order β, as shown in Fig. 3, is the generalization of VDTA based integer order inductor [10] in fractional domain.

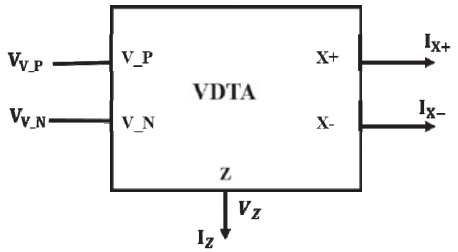


Fig. 1. Circuit symbol for VDTA block

The routine circuit analysis results in the inductance value of the FI as presented in (6).

$$L_{\beta} = \frac{C_{\alpha}}{g_{m1} g_{m2}} \tag{6}$$

where C_{α} represents FC of order α and is designed using RC Domino ladder network truncated to 11-12 blocks.

III. Applications Of Proposed FI

The proposed FI is functionally verified by using it as a circuit element in three fractional order $RL_{\beta}C_{\alpha}$ filters namely FLPF, FHPF and FBPF. The filters are designed by lifting the appropriate node of a parallel resonator. The detailed mathematical analysis has been presented in following subsections.

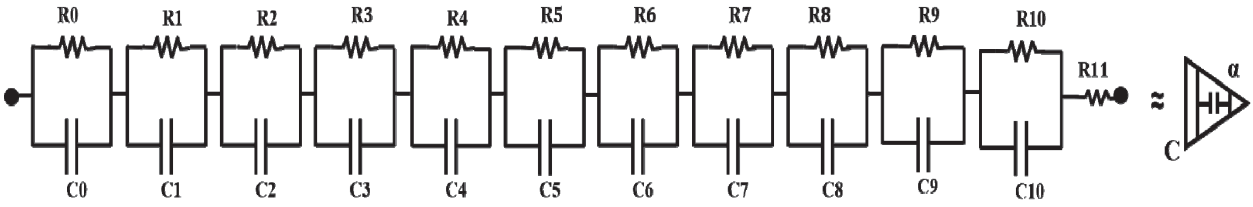


Fig. 2. Domino ladder based fractional order capacitor

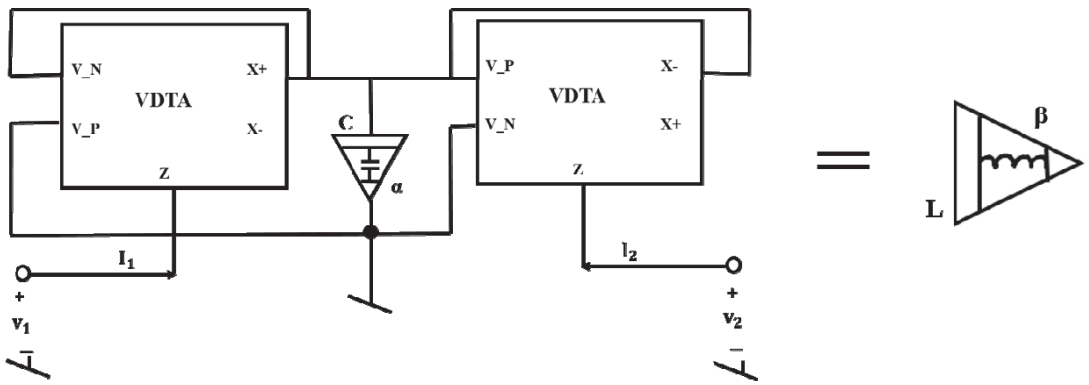


Fig. 3. Proposed fractional order inductance using VDTA and FC

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.

A. Fractional Band pass Filter

The circuit diagram of the designed FBPF with C = C_α and L = L_β is shown in Fig. 4. The transfer function (TF) of this FBPF is computed as

$$H(s)=\frac{V_{out}}{V_{in}}=\frac{\frac{s^\beta}{RC_\alpha}}{s^{\alpha+\beta}+\frac{s^\beta}{RC_\alpha}+\frac{1}{L_\beta C_\alpha}} \tag{7}$$

The magnitude of the TF is calculated as

$$\left|H(j\omega)\right|=\left|\frac{\frac{e^{j\pi\beta/2}\omega^\beta}{RC_\alpha}}{e^{j\pi(\alpha+\beta)/2}\omega^{\alpha+\beta}+\frac{e^{j\pi\beta/2}\omega^\beta}{RC_\alpha}+\frac{1}{L_\beta C_\alpha}}\right| \tag{8}$$

Equating the magnitude to 1/√2 times the maximum gain, the half power frequencies (f_L, f_H) are calculated. Thus the generalized equation for determining the cut-off frequencies is presented in (9).

$$\begin{aligned}&\omega^{2(\alpha+\beta)}+\frac{2}{RC_\alpha}cos(\frac{\alpha\pi}{2})\omega^{\alpha+2\beta}\\&+\frac{2}{L_\beta C_\alpha}cos\left[\frac{(\alpha+\beta)\pi}{2}\right]\omega^{\alpha+\beta}\\&-\frac{1}{R^2C_\alpha{}^2}\omega^{2\beta}+\frac{2}{RL_\beta C_\alpha{}^2}cos(\frac{\beta\pi}{2})\omega^\beta+\frac{1}{L_\beta{}^2C_\alpha{}^2}=0\end{aligned} \tag{9}$$

The center frequency (f_c) corresponds to the maximum gain frequency.

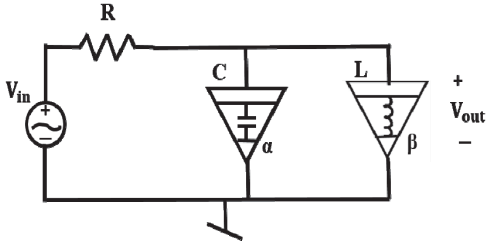


Fig. 4. The proposed FI based FBPF

B. Fractional Low pass Filter

The proposed FI based FLPF is shown in Fig. 5. The TF of FLPF is evaluated as follows:

$$H(s)=\frac{V_{out}}{V_{in}}=\frac{\frac{1}{L_\beta C_\alpha}}{s^{\alpha+\beta}+\frac{s^\beta}{RC_\alpha}+\frac{1}{L_\beta C_\alpha}} \tag{10}$$

The magnitude of the transfer function is calculated as

$$\left|H(j\omega)\right|=\left|\frac{\frac{1}{L_\beta C_\alpha}}{e^{j\pi(\alpha+\beta)/2}\omega^{\alpha+\beta}+\frac{e^{j\pi\beta/2}\omega^\beta}{RC_\alpha}+\frac{1}{L_\beta C_\alpha}}\right| \tag{11}$$

Equating the magnitude to 1/√2 times the DC gain, the cutoff frequency can be computed. On simplification, the generalized equation is obtained as below.

$$\begin{aligned}&\omega^{2(\alpha+\beta)}+\frac{2}{RC_\alpha}cos(\frac{\alpha\pi}{2})\omega^{\alpha+2\beta}\\&+\frac{2}{L_\beta C_\alpha}cos\left[\frac{(\alpha+\beta)\pi}{2}\right]\omega^{\alpha+\beta}\\&+\frac{1}{R^2C_\alpha{}^2}\omega^{2\beta}+\frac{2}{RL_\beta C_\alpha{}^2}cos(\frac{\beta\pi}{2})\omega^\beta-\frac{1}{L_\beta{}^2C_\alpha{}^2}=0\end{aligned} \tag{12}$$

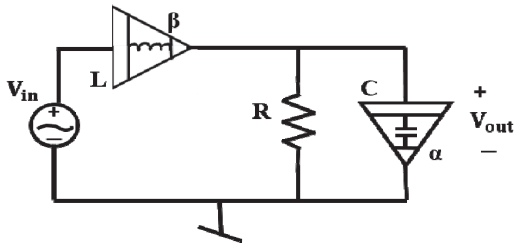


Fig. 5. The proposed FI based FLPF

C. Fractional High pass Filter

The circuit of proposed FHPF is given in Fig. 6 and its TF is expressed as

$$H(s)=\frac{V_{out}}{V_{in}}=\frac{s^{\alpha+\beta}}{s^{\alpha+\beta}+\frac{s^\alpha R}{L_\beta}+\frac{1}{L_\beta C_\alpha}} \tag{13}$$

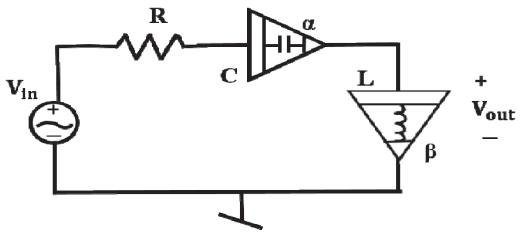


Fig. 6. The proposed FI based FHPF

The magnitude of the TF of (13) is calculated as

$$\left|H(j\omega)\right|=\left|\frac{\omega^{\alpha+\beta}e^{j\pi(\alpha+\beta)/2}}{e^{j\pi(\alpha+\beta)/2}\omega^{\alpha+\beta}+\frac{e^{j\pi\alpha/2}\omega^\alpha R}{L_\beta}+\frac{1}{L_\beta C_\alpha}}\right| \tag{14}$$

Equating the magnitude response to 1/√2 times the maximum gain the half power frequency is calculated. On simplification, the following generalized equation is obtained.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.

$$\omega^{2(\alpha+\beta)} - \frac{2R}{L_\beta} cos(\frac{\beta\pi}{2})\omega^{2\alpha+\beta} - \frac{2}{L_\beta C_\alpha} cos\left[\frac{(\alpha+\beta)\pi}{2}\right]\omega^{\alpha+\beta} - \frac{R^2}{L_\beta{}^2}\omega^{2\alpha} - \frac{2R}{C_\alpha L_\beta{}^2} cos(\frac{\alpha\pi}{2})\omega^{\alpha} - \frac{1}{L_\beta{}^2 C_\alpha{}^2} = 0 \quad (15)$$

IV. Simulation Results

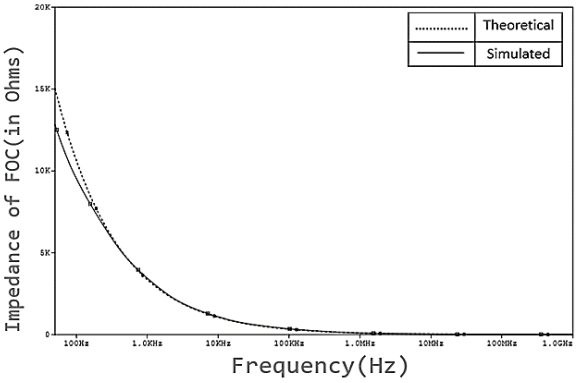
The SPICE simulations are carried out for functionality assessment of proposed FI. The VDTA [9] is biased using DC power supplies of ± 0.9V and 180 nm TSMC technology node is used as process parameters. The channel length for all transistors is chosen as 0.36 µm. The width of NMOS transistors is considered to be 3.6 µm and that for PMOS devices is taken as 16.64 µm. All biasing current sources are taken as 150 µA and the transconductances are obtained as g_{m1} = 605.042 µA /V and g_{m2} = 594.205 µA /V. The FC of Fig. 2 [2] is designed for α = 0.5 and C_α = 3.75 µ℧/$s^{0.5}$. The ladder wise component values are chosen as R_0 = 330 kΩ, C_0 = 4.7 µF, R_1 = 82 kΩ, C_1 = 3.1 µF, R_2 = 33 kΩ, C_2 = 1 µF, R_3 = 12 kΩ, C_3 = 470 nF, R_4 = 4.7 kΩ, C_4 = 168 nF, R_5 =2kΩ, C_5 = 68 nF, R_6 = 736 Ω, C_6 = 27 nF, R_7 = 270 Ω, C_7 = 10 nF, R_8 = 120 Ω, C_8 = 4.7 nF, R_9 = 47 Ω, C_9 = 1 nF, R_{10} = 8.2 Ω, C_{10} = 2.2 nF and R_{11} = 18.2 Ω.

The simulated FC impedance magnitude plot is shown in Fig. 7(a) whereas the FC value is plotted in Fig. 7(b). The phase responses of FC are shown in Fig.7 (c). These responses have been observed to be quite identical to the theoretical responses i.e. C_α = 3.75 µ℧/$s^{0.5}$ and $\angle Z_C(j\omega) = \theta_{C\alpha}$ = - 45°.

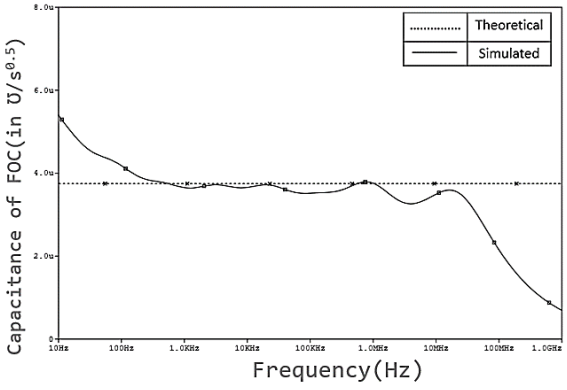
The simulated impedance magnitude plot of floating FI is depicted in Fig. 8 (a) whereas the variation of FI inductance value with frequency is plotted in Fig. 8 (b) and impedance phase response is shown in Fig. 8 (c). It may be observed from Fig. 8 (a) that the impedance increases with the increase in frequency which confirms to the expression for impedance relation of floating FI in a frequency range of few hundred Hz to 1MHz. Corresponding to the same range of operational frequency a nearly constant value of inductance L_β = 10.435 Ω/$s^{0.5}$ varying well within ±10% of the relative error. Further, the phase response shows a similar trend where it yields a nearly constant phase which matches the theoretically obtained phase $\angle Z_L(j\omega) = \theta_{L\beta}$ = + 45°.

The proposed filter applications are simulated next to demonstrate the applicability of the proposed floating FI. The magnitude responses of the transfer function have been analysed for the 3 dB cut-off frequencies corresponding to the design specifications of each filter.

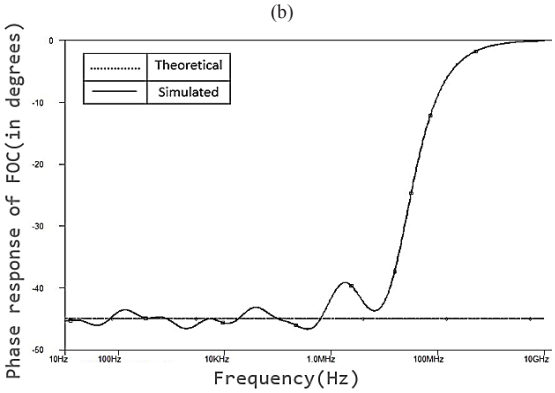
The FBPF employs an L_β - C_α network in series with R where R=1 kΩ, C_α = 3.75 µ℧/$s^{0.5}$ and L_β = 10.435 Ω/$s^{0.5}$. This filter design yields a center frequency f_c = 1.2598 kHz and BW = f_H - f_L = 7.786 kHz as obtained from the corresponding magnitude response plot of Fig. 9.



(a)



(b)



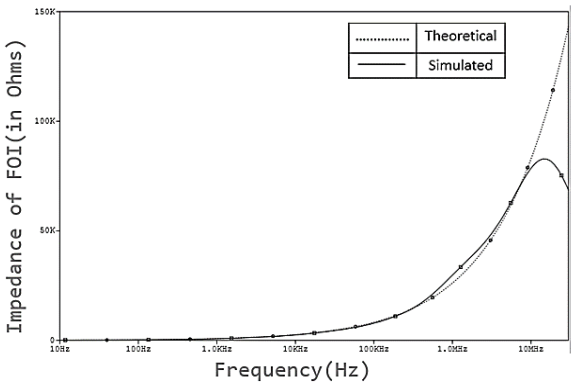
(c)

Fig. 7. (a) Impedance provided by the fractional order capacitance; (b) variation of fractional order capacitance with frequency; (c) phase response of fractional order capacitance

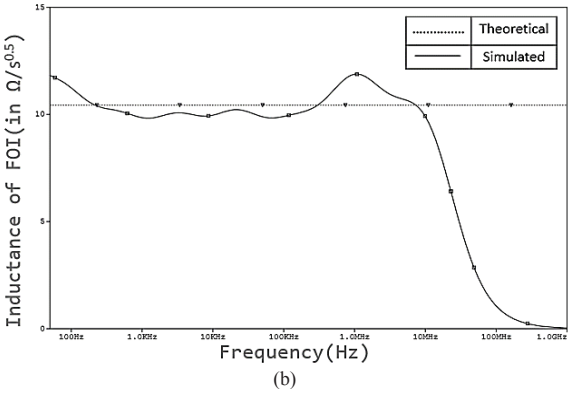
The FLPF employs a parallel R- C_α network in series with L_β where R = 5 kΩ, C_α = 3.75 µ℧/$s^{0.5}$ and L_β = 10.435 Ω/$s^{0.5}$. The magnitude response is shown in Fig. 10 and yields a 3dB frequency of 4.2695 kHz.

Similarly, the fractional order High pass filter employs a series $RL_\beta C_\alpha$ network where R= 3kΩ, C_α = 3.75 µ℧/$s^{0.5}$ and L_β = 10.435 Ω/$s^{0.5}$. The cut-off frequency from the corresponding magnitude plot of Fig. 11 is obtained as 12.516 kHz. Thus, the usage of a fractional order inductor yield the appropriate responses with the added advantage of obtaining a roll-off which is non-integral multiple of the standard 20dB/decade slope.

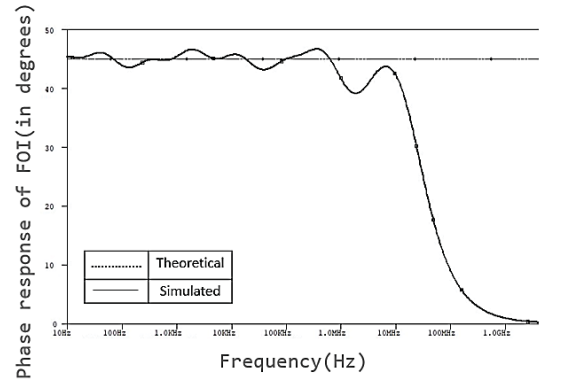
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.



(a)



(b)



(c)

Fig. 8. (a) Impedance provided by the fractional order inductance; (b) variation of fractional order inductance with frequency; (c) phase response of fractional order inductance

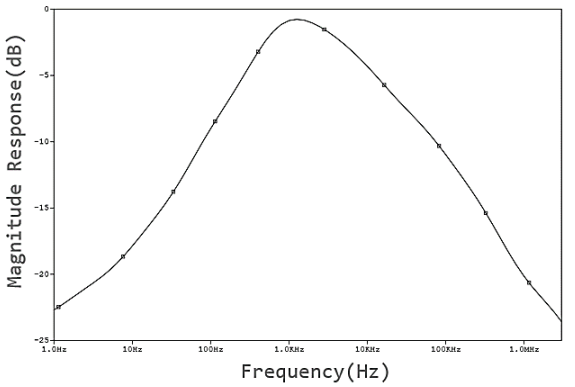


Fig. 9. The FBPF magnitude response

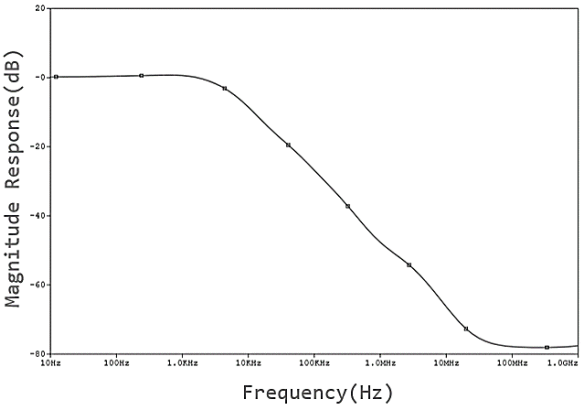


Fig. 10. The FLPF magnitude response

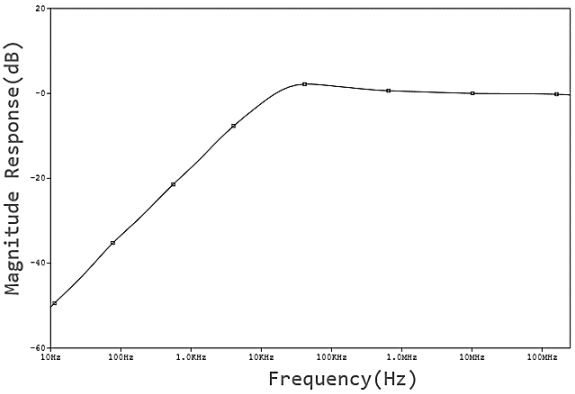


Fig. 11. The FHPF magnitude response

V. CONCLUSION

In this paper, a floating FI designed using VDTA has been presented. The proposed design is verified with the use of SPICE simulations for fractional order $\beta = 0.5$. The behavior of the proposed FI was tested by comparing the theoretical impedance magnitude and phase responses with the simulated ones. Further, proposed FI is employed in three different filter applications.

The proposed fractional inductor, when compared to other existing implementations in literature has an edge in terms of the number of additional passive components employed and the simplicity of implementation of the active device.

REFERENCES

[1] A. G. Radwan and K. N. Salama, "Passive and Active Elements Using Fractional $L_\beta C_\alpha$ Circuit," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 58, no. 10, pp. 2388-2397, Oct. 2011.

[2] R. Verma, N. Pandey and R. Pandey, "Realization of a higher fractional order element based on novel OTA based IIMC and its application in filter," Analog Integr Circ Sig Process, vol. 97, no. 1, pp. 177-191, Oct. 2018.

[3] N. Herencsar, "Balanced-Output CCCFOA and its Utilization in Grounded Inductance Simulator with Various Orders," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, pp. 1-5, 2018.

[4] A. Soltan, A. G. Radwan, and A. M. Soliman, "CCII based fractional filters of different orders," J. Adv. Res., vol. 5, no. 2, pp. 157–164,2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.

[5] D. Susan and S. Jayalalitha, "Frequency selective circuits using Op-Amp based simulated inductor," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, pp. 2310-2314, 2017.

[6] G. Tsirimokou, C. Psychalinos, A. S. Elwakil and K. N. Salama, "Electronically Tunable Fully Integrated Fractional-Order Resonator," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 2, pp. 166-170, Feb. 2018.

[7] K. H. Khattab, A. H. Madian and A. G. Radwan, "CFOA-based fractional order simulated inductor," 2016 IEEE 59th International Midwest Symposium on Circuits and Systems (MWSCAS), Abu Dhabi, , pp. 1-4, 2016.

[8] A. Adhikary, P. Sen, S. Sen and K. Biswas, "Design and Performance Study of Dynamic Fractors in Any of the Four Quadrants," in Circuits, Systems, and Signal Processing , vol. 35, no. 6, pp. 1909–1932, Jun. 2016.

[9] Y. Abdullah, F. Kaçar and H. Kuntman, "New Simple CMOS Realization of Voltage Differencing Transconductance Amplifier and Its RF Filter Application," Radioengineering, Vol. 20, Sep. 2011

[10] D. Prasad and D. Bhaskar, "Grounded and Floating Inductance Simulation Circuits Using VDTAs," Circuits and Systems, Vol. 3, No. 4, pp. 342-347, 2012.

[11] A. Kartci, N. Herencsar, L. b. Brancik and K. N. Salama, "CMOS-RC Colpitts Oscillator Design Using Floating Fractional-Order Inductance Simulator," 2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), Windsor, ON, Canada, pp. 905-908, 2018.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 09,2020 at 09:04:53 UTC from IEEE Xplore. Restrictions apply.

2020 6th International Conference on Advanced Computing & Communication Systems (ICACCS)

YOLO v3-Tiny: Object Detection and Recognition using one stage improved model

[1]Pranav Adarsh
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pranavadarsh95@gmail.com

[2]Pratibha Rathi
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
pratibharathi96@gmail.com

[3]Manoj Kumar
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
mkumarg@dce.ac.in

***Abstract*—Object detection has seen many changes in algorithms to improve performance both on speed and accuracy. By the continuous effort of so many researchers, deep learning algorithms are growing rapidly with an improved object detection performance. Various popular applications like pedestrian detection, medical imaging, robotics, self-driving cars, face detection, etc. reduces the efforts of humans in many areas. Due to the vast field and various state-of-the-art algorithms, it is a tedious task to cover all at once. This paper presents the fundamental overview of object detection methods by including two classes of object detectors. In two stage detector covered algorithms are RCNN, Fast RCNN, and Faster RCNN, whereas in one stage detector YOLO v1, v2, v3, and SSD are covered. Two stage detectors focus more on accuracy, whereas the primary concern of one stage detectors is speed. We will explain an improved YOLO version called YOLO v3-Tiny, and then its comparison with previous methods for detection and recognition of object is described graphically.**

Keywords—Computer vision; YOLO v3; Faster RCNN; Deep learning; YOLO v3-Tiny; Object detection; image processing; Convolutional Neural Networks.

I. INTRODUCTION

In the recent few years, diverse research work happened to develop a practical approach to accelerate the development of deep learning methods. Numerous developments accomplished excellent results and followed by continuous reformations in deep learning procedures. Object localization is the identification of all the visuals in a photograph, incorporating the precise location of those visuals. By using deep learning techniques [1] [2] for object identification and localization, computer vision has reached a new zenith. Due to significant inconstancies in viewpoints, postures, dimensions, and lighting positions, it is challenging to succeed in the identification of objects perfectly. Accordingly, considerable concern has been given by researchers to this area in the past few years. There are two types of object detection algorithms. Object detection algorithms using region proposal includes RCNN [3], Fast RCNN [4], and Faster RCNN [5], etc. These techniques create region proposal networks (RPN) [6], and then the region proposals are divided into categories afterward. On the other side, object detection algorithms using regression includes SSD [7] and YOLO [8], etc. These methods also generate region proposal networks (RPN) but divide these region proposals into categories at the moment of generation. All of the procedures mentioned above have significant accomplishments in object localization and recognition. YOLO consolidates labels in diverse datasets to form a tree-like arrangement, but the merged labels are not reciprocally exclusive. YOLO9000 [9] enhances YOLO to recognize targets above 9000 categories employing hierarchical arrangement. Whereas YOLOv3 [10] uses multilabel classification, it replaces the approach of estimating the cost function and further exhibits meaningful improvement in distinguishing small targets. The arrangement of this paper is as follows. Below in section 2, background information of object detection methods is covered. It includes two stage detectors with their methodologies and drawbacks. Section 3 elaborates one stage detectors and the improved version YOLO v3-Tiny [10] [11]. Section 4 describes implementation results and comparison of object detection methods based on speed and accuracy. Finally, section 5 summarizes the conclusion.

II. BACKGROUND

In this section, we present background information. It elaborates the most representative and pioneering two-stage object detection methods with their significant contributions in object detection. First, we examine their methodologies and then explain their drawbacks.

A. HOG

HOG is a feature descriptor that is extensively applied in various domains to distinguish objects by identifying their

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

shapes and structures. Local object structure, pattern, aspect, and representation can usually be characterized by the arrangement of gradients of local intensity or the ways of edges. In the HOG [12] detection method, the first step is to break the source image into blocks and then distribute each block in small regions. These regions are called cells. Commonly, the blocks of image overlap each other, due to this corresponding cell may be a part of many blocks. For each pixel inside the cell, it calculates the gradients vertically and horizontally.

Drawbacks of HOG method- Due to the emergence of deep learning and its significant applications, the reasonable opinion was to displace classifiers deployed on HOG [12] methodology with a classifier based on convolutional neural network [2][13] because of its comparatively higher accuracy. But there were some problems. The computation cost of convolutional neural networks was high, and the speed is too slow. Therefore, it was challenging to run CNN based classifier on numerous patches produced by the detection approach of sliding window. This difficulty was resolved by RCNN [3]. It employs an algorithm based on object proposals termed selective search method [14]. This approach decreases the number of bounding boxes to 2000 region proposals that were supplied to the R-CNN classifier.

B. RCNN

Region based convolutional neural networks (RCNN) algorithm [3] uses a group of boxes for the picture and then analyses in each box if either of the boxes holds a target. It employs the method of selective search to pick those sections from the picture. In an object, the four regions are used. These are varying scales, colours, textures, and enclosure.

Drawbacks of RCNN method- Based on a selective search [14], 2,000 sections are excerpted per image. For every region or part of the image, we have to select features using CNN. For this, if we have 'i' number of images, then selected regions will become i×2,000. The whole method of target identification through RCNN utilizes the following three models: Linear SVM classifier for the identification of objects, CNN is employed for characteristic extraction, and a regression model is required to tighten the bounding boxes. All these three processes combine to take a considerable amount of time. It increases the running time of RCNN method. Therefore, RCNN needs almost 40 to 50 seconds to predict the result for several new images [15].

C. FAST RCNN

In place of using three different models of RCNN, Fast RCNN [4] employs one model to excerpt characteristics from the different regions. Then it distributes the regions into several categories based on excerpted features, and the boundary boxes of recognized divisions return together. Fast RCNN uses the method of spatial pyramid pooling [16] to calculate only one CNN representation for the whole image. It passes one region for each picture to a particular convolutional network model by replacing three distinct models for excerption of characteristics, distributing into divisions, and producing bounding boxes.

Drawbacks of Fast RCNN method- Fast RCNN also employ a selective search method [14] to detect concerned regions. This method is prolonged and demands a lot of time. Usually, for the detection of objects, this complete procedure needs almost two seconds for each picture. Therefore its speed is quite good in contrast to RCNN [15]. However, if we contemplate extensive real-life datasets, then the execution of fast RCNN approach is still lacked in speed.

D. FASTER RCNN

Faster RCNN [5] is a transformed variant of fast RCNN. The significant difference between both is that faster RCNN implements region proposal network (RPN) [6] [17], but fast RCNN utilizes a selective search technique for producing concerned regions. In input, RPN accepts feature maps of picture and produces a collection of object recommendations and an objectness score per recommendation in output. Usually, this approach takes ten times less time in contrast to fast RCNN approach because of RPN.

Drawbacks of faster RCNN method- To excerpt all the targets in a given picture, this procedure needs multiple passes for that particular picture. Different systems are working in a sequence therefore, the performance of the upcoming operation is based on the performance of preceding operations. This approach uses region proposal networks to localize and identify the objects in a picture. But RPNs do not contemplate the complete picture because it uses only those portions of the picture, which have high probabilities of the presence of targets [15].

Table 1: Comparison of two stage object detectors

Type	Characteristics	Computation time	Drawbacks
RCNN	To generate regions, it uses selective search. From each picture, it extracts around 2000 regions.	Forty to Fifty seconds	Time taken for prediction is large because several regions pass through CNN definitely, and it employs three distinct models for the detection of targets.
Fast RCNN	To excerpt the features, each picture passes one time through CNN. All distinct models applied in RCNN are combined collectively to	Two seconds	The method used is prolonged and time-consuming Therefore, computation time is still high.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

	form a single model. It employs selective search method on the feature maps to produce result for target recognition.		
Faster RCNN	The previous approach is replaced with the region proposal networks. Therefore, this procedure works much faster compared to previous methods.	0.2 seconds	Object region proposal is time-consuming. Different types of systems are operating in sequence. Thus, the performance of entire procedure is based on the working of the preceding operations.

III. ONE STAGE DETECTORS

Two stage detectors provide adequate accuracy, but the time taken for computation is high. Therefore, to process in less time by managing sufficient accuracy, one stage detectors are proposed. Some algorithms in one stage model are SSD and the variants of YOLO. By improving the architecture of two stage models and introducing some changes such as eliminating pipeline, one stage model has achieved excellent speed. But, it has not attained sufficient accuracy at the same time. Hence, continuous changes are under process by researchers.

A. YOLO v1

By using the approach of one stage detector YOLO, an input picture is distributed to a system of S×S grids [8]. The detection of the object depends on every grid of the input image. Grid cells are employed to predict targets inside boundary boxes. Five parameters are predicted for every boundary box. These five elements are i, j, k, l, and c. In the input picture, the centre of target inside the box is denoted by 'i' and 'j' coordinates. Here, 'k', 'l', and 'c' represent height, width, and score for confidence respectively. 'c' is measured as the probability of containing the target inside boundary box.

YOLO v1 uses the Darknet framework and ImageNet-1000 dataset to train the model. It distributes the given picture to a grid of S×S cells. For every cell in the network, it computes confidence for 'n' bounding boxes. The predicted result is encoded into a tensor as $S \times S \times (n \times 5+ p)$ [8]. Here, input image is divided into total S×S sub-images. In n×5, five

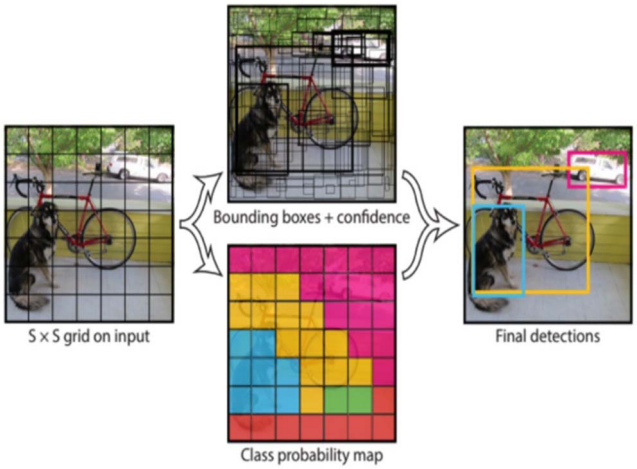


Fig.1: S×S grid representation of YOLO [8]

represents the detection of five attributes for each bounding box, which are height, weight, confidence score, and centre coordinates(x, y) of detected objects. Here, 'p' represents the probability of class. YOLO v1 also has many limitations. Therefore, the use of YOLO v1 is restricted to some extent. Limitations of YOLO version1 are based on the closeness of the objects in the picture. If the objects appear as a cluster, they could not find the small objects. If the dimensions of the object are different from the image used in training data, then this architecture found difficulty in the localization and detection of objects [15][18]. The primary concern is to locate objects in a given picture due to the error of localization.

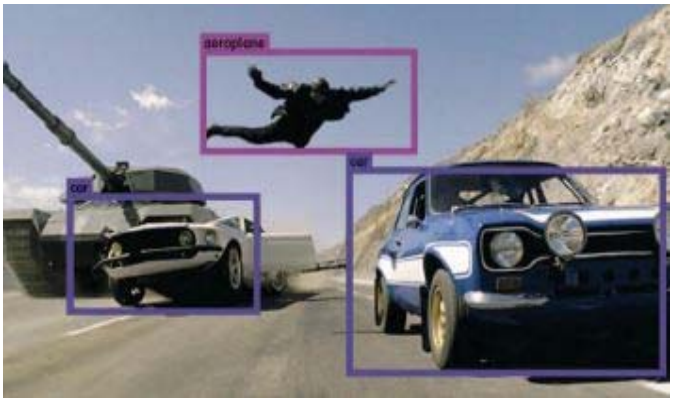


Fig.2: Drawback of Yolo v1 [8]

YOLO v1 fails sometimes, as in the above image it recognizes the man as an airplane.

B. YOLO v2

Yolo v2 supersedes Yolo by offering a great balance between running time and accuracy. For better accuracy, Yolo v2 introduces batch normalization, which helps to enhance 2 percent in map by attaching it into each layer of convolution. High resolution classifier is used to operate thoroughly by modifying its filters for giving a more extensive understanding of network time to work excellently. When it comes to the prophecy of bounding boxes, then by eliminating entirely

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

connected layers to anchor boxes makes decrement of map value by 0.3, but recall value is incremented by 7% [9]. Hence, it gives more potential to detect the day to day objects. Fine grained features help to identify tiny objects by reshaping the layers employing a modified approach called pass through. As we have achieved accuracy now, our goal is to maintain a balance of running time with it. For this Yolo v2 uses darknet 19, although it could use google net, but that reduces accuracy by 2%. For learning to locate targets from any visual, Yolo 9000 is employed across nine thousand categories with a 9418 node WordTree. It is certainly progress for concluding the localization and distribution.

C. YOLO v3

Object detection is used in many fields of human life, for example, health and education, etc. As all these fields are growing rapidly, so to match their requirements, one stage models also need improvement. The next advanced variant of YOLO is version 3 that uses logistic regression to compute the targetness score. It gives the score for all targets in each boundry box. YOLO v3 can give the multilabel classification because it uses a logistic classifier for each class in place of the softmax layer used in YOLO v2. YOLO v3 uses darknet 53. It has fifty-three layers of convolution. These layers are more in-depth compared to darknet 19 used in YOLO v2. Darknet-53 contains mainly 3x3 and 1x1 filters along with bypass links [8] [9] [10]. The formulas given below explain the transformation of network output for obtaining bounding box predictions. Here, d_x and d_y are center coordinates, d_w is width and d_h is the height of predicted result. Top left coordinates of the grid are m_x and m_y. Network outputs are t_x, t_y, t_w and t_h. Anchor dimensions for the box are p_h and p_w.

$$d_x = \sigma(t_x) + m_x \quad (1)$$

$$d_y = \sigma(t_y) + m_y \quad (2)$$

$$d_w = p_w e^{t_w} \quad (3)$$

$$d_h = p_h e^{t_h} \quad (4)$$

By using threshold value, a filter is applied to remove the box having class score less than the threshold value chosen. Because a score with less value represents that the box is insufficient for identifying the classes. Even after filtering by using a threshold value for the class scores, a lot of overlapping boxes still remain. When many boxes are overlapping with each other, then select only one box out of those overlapping boxes and identify the object. So, the second filter is used for choosing the desired boxes, which termed as nonmaximum suppression (NMS) [19]. It uses intersection over union (IOU) function.

$$IOU = \frac{B_1 \cap B_2}{B_1 \cup B_2} \quad (5)$$

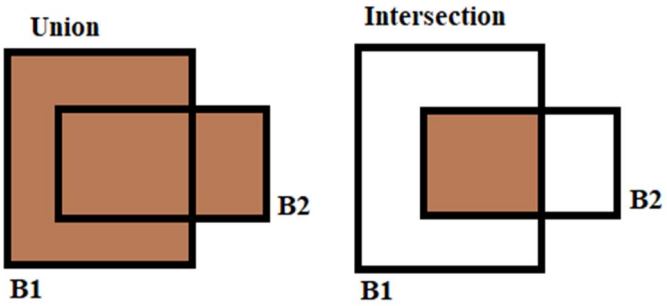


Fig. 3: Illustration depicting the definition of union and intersection

Above-left corner (a1, b1) and below right corner (a2, b2) are used to determine a box. For calculating the area of the rectangle, multiply its height (b2-b1) and its width (a2-a1). Then the coordinates (ai1, bi1, ai2, bi2) for the intersection of 2 boxes are obtained. Here, ai1 and bi1 are the highest value of a1 and b1 coordinate-position of the two boxes. Similarly, ai2 and bi2 is the lowest value of the a2 and b2 coordinate-position of the 2 boxes.

$$\text{union}(A, B) = A + B - \text{inter}(A, B) \quad (6)$$

$$\text{inter_area} = (ai2 - ai1) \times (bi2 - bi1) \quad (7)$$

$$\text{box1_area} = (box1[3] - box1[1]) \times (box1[2] - box1[0]) \quad (8)$$

$$\text{box2_area} = (box2[3] - box2[1]) \times (box2[2] - box2[0]) \quad (9)$$

$$\text{union_area} = (\text{box1_area} + \text{box2_area}) - \text{inter_area} \quad (10)$$

$$\text{IOU} = \text{inter_area}/\text{union_area} \quad (11)$$

The advantage of YOLO v3 over YOLO v2 is that some changes are included in error function and for objects of small to a considerable size detection occurs on three scales. The multiclass problem turned in a multilabel problem, and the performance improved over small size objects.

D. SSD

SSD is a single shot detector [7] [20]. It manages an excellent balance of speed with the accuracy of result. In this, we apply for single time a CNN based model to the input picture for computing the feature map. It also employs anchor boxes similar to faster RCNN at various aspect ratios and learns the offset instead of determining the box. The processing occurs on numerous layers of CNN, where every layer functions on a varying range of scale and uses multiple feature maps. Therefore, the detection of targets of several sizes is possible. Experimentally, SSD has much better accuracy on different datasets even on inputs pictures of small size as compared to the other single stage methods.

E. SSD vs YOLO

Unlike YOLO, SSD does not divide the image into grids of random size. For every location of the feature map, it predicts the offset of predefined anchor boxes (default boxes). Relative to the corresponding cell, each box has a fixed size, proportion, and position. In a convolutional manner, all the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

anchor boxes cover the entire feature map. Anchors of SSD are slightly different from the anchors of YOLO. Because YOLO makes all the predictions from a single grid, and the size of anchors used by YOLO ranges from dimensions of one grid cell to the dimensions of the entire picture. The anchors of SSD specialize its detector for distinct feasible viewpoints and dimensional ratios of its target shapes, but not enough on the size of targets. The calculation for the anchors of SSD uses a simple formula, while the anchors of YOLO are calculated by applying k-means clustering on the training data [21] [15]. SSD doesn't use confidence score, but YOLO calculates it to show the faith in predicted results. A unique background class is employed by SSD for this work. A low value of confidence score in YOLO is equivalent to the predicted output of background class in SSD. Both indicate that for the detector, the possibility of getting a target is null [18].

F. YOLO v3-Tiny

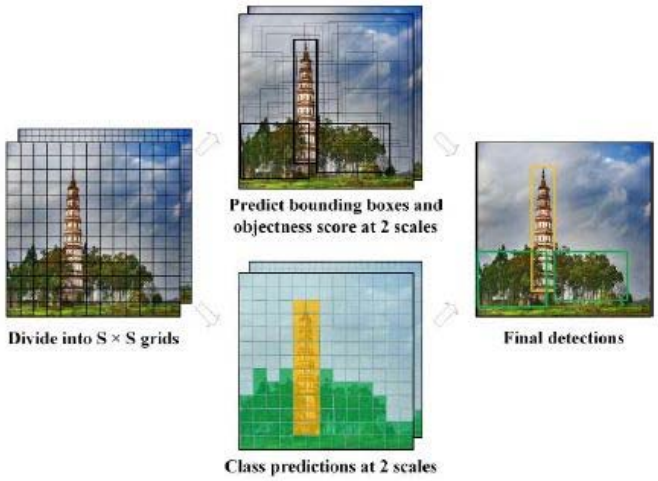


Fig. 4: Main process of YOLO v3-Tiny [11]

YOLO v3, with the decreased depth of the convolutional layer, is another version called YOLO v3-Tiny. It was proposed by Joseph Redmon [10]. Therefore, the running speed is significantly increased (approximately 442% faster than the former variants of YOLO), but detection accuracy is reduced. Darknet-53 architecture of YOLO v3 employs several 1x1 convolution layers along with 3x3 convolution layers for extracting features [22]. YOLO v3-Tiny uses pooling layer and reduces the figure for convolution layer. It predicts a three-dimensional tensor that contains objectness score, bounding box, and class predictions at two different scales. It divides a picture into S×S grid cells. For final detections, we will ignore the bounding boxes for which the objectness score is not best. For extracting features, convolution layers and max-pooling layers are utilized in the feed forward arrangement of YOLO v3-Tiny [11] [23] [24]. Prediction of bounding boxes occurs at two different feature map scales, which are 13×13, and 26×26 merged with an upsampled 13×13 feature map.

Table 2: Performance comparison of YOLO with their Tiny versions[25]

Detector	Number of n-layers	FLOPS	FPS	map value	Used set of data
YOLO v1	26.00	Not-given	45.00	63.50	VOC-data
YOLO v1-Tiny	9.00	Not-given	155.00	52.80	VOC-data
YOLO v2	32.00	62.95	40.00	48.20	COCO-data
YOLO v2-Tiny	16.00	05.42	244.00	23.60	COCO-data
YOLO v3	106.00	140.70	20.00	57.80	COCO-data
YOLO v3-Tiny	24.00	05.57	220.00	33.20	COCO-data

In Table 2, we can see that on different datasets, the tiny versions of YOLO have a lower value for map as compared to their original versions.

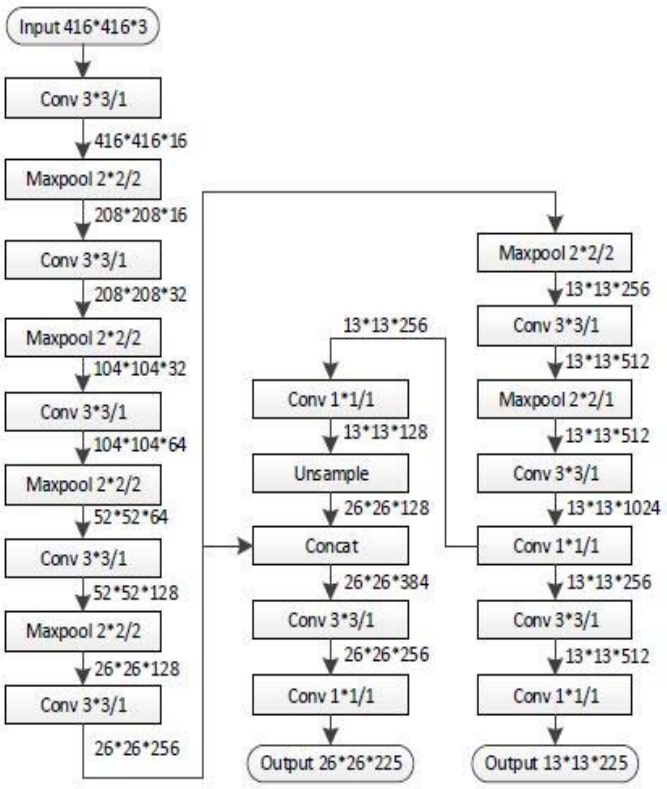


Fig.5: Architecture of YOLO v3-Tiny [11]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

IV. IMPLEMENTATION RESULTS AND ANALYSIS

It is tough to make a clear comparison between different object detection methods. So, we can't give a straight decision on the best model. For many real-life applications, we make choices to create an equilibrium of accuracy with speed. Therefore, we need to be aware of other characteristics that have a significant impact on performance. For example, matching strategy and IOU threshold, ratio of positive anchor and negative anchor, training dataset, utilization of varying scale and cropped pictures for training, location loss function, pace of learning, and learning rate decay etc.

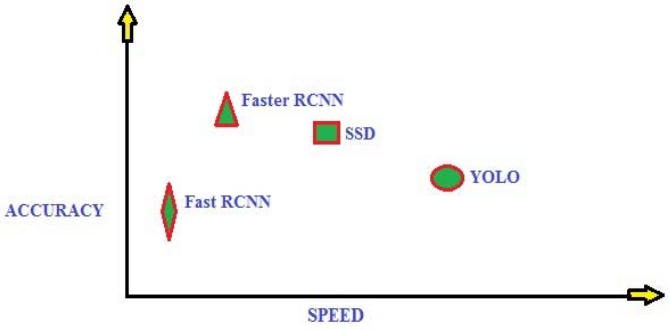


Fig. 6: Speed vs. Accuracy for object detection methods

We have thoroughly examined the characteristics of three procedures for object detection, which are faster RCNN, YOLO v3, and SSD [26]. If the requirement is accuracy (excellent quality of correctness), then faster RCNN is the best as it is much accurate compared to SSD and YOLO v3. But if accuracy is not the primary concern, we want super-fast speed, YOLO v3 takes less time compared to SSD and faster RCNN. But if at the same time requirement is of excellent accuracy and less running time, then SSD is a more favorable recommendation, as its speed is better than faster RCNN, and its accuracy is better compared to YOLO v3.



Fig.7: Accuracy comparison for different sizes of target objects

The chart given above in Fig. 7 represents the faster RCNN, YOLO, and SSD performance on targets of distinct sizes. In accuracy comparison for the big targets, SSD performs alike faster RCNN. But as the object size decreases, the gap increases [15] [18] [27].

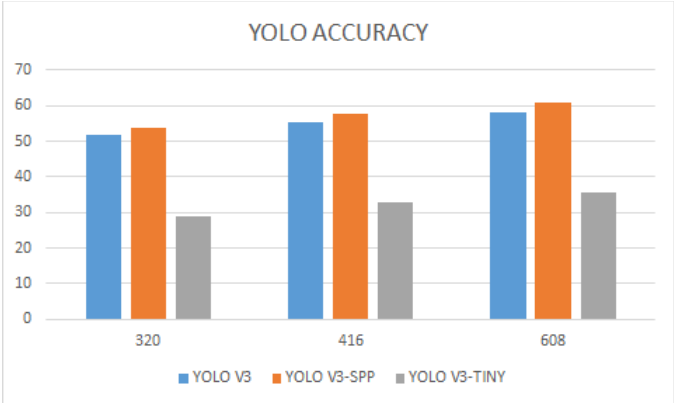


Fig.8: Accuracy comparison of YOLO algorithms

YOLO v3-Tiny is a lightweight variant of YOLO v3, which takes less running time and less accuracy when examined with YOLO v3. In Fig. 8, we compared the accuracy of different versions of YOLO algorithms for given image pixels.

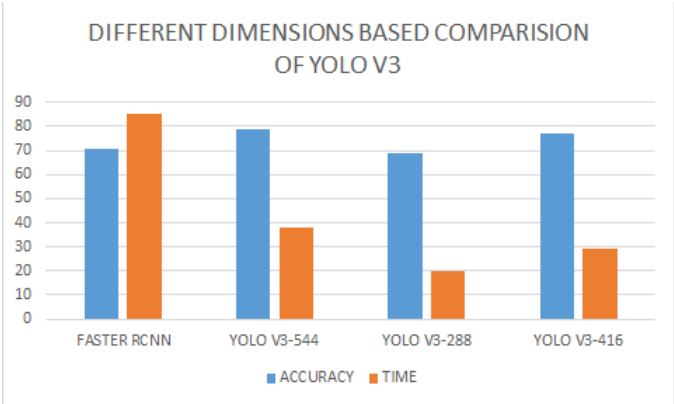


Fig.9: Accuracy and speed comparison of YOLO with faster RCNN

In Fig. 9, the accuracy and speed performance of YOLO algorithms is compared with faster RCNN [27]. As per the result, we can say that YOLO v3 takes less time compared to Faster R-CNN, and YOLO v3-tiny is faster than YOLO v3.

We have implemented YOLO v3 and YOLO v3-tiny using conda, open CV python, and TensorFlow 2.0 with the darknet 53 model, which is earlier trained on the imagenet dataset for classification. As training a model from scratch is much complicated, so we employed transfer learning on our model using VOC 2012 dataset for localization of objects. We have used a PC comprising Windows 10, intel core i5 eighth-generation quad-core CPU, NVIDIA Geforce GTX 1050 GPU, and a RAM of size 8 gigabytes. It also demands an internet of extremely high-speed. Programming Language used is python. We have taken input as preprocessed image. The output is in the form of bounding boxes, accuracy score percentage, and class label. During the training of model, ninety percent of data is applied for training, and the rest ten percent is used for the validation of our model. We trained our model for localizing twenty classes that include person, dog, and car, etc. During batch normalization, the size of each batch is sixteen and requires approximately twenty iterations

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

to get a valid result. The duration of the period needed for each iteration is around ten minutes. Then the trained models of YOLO v3 and YOLO v3-Tiny are applied to the pictures taken by us in our college campus to examine the execution time and accuracy (map value) of the result obtained. Speed is calculated in terms of frame per second, which is the number of images processed by the model in each second. Mean average precision is computed to estimate accuracy. After each iteration, precision and recall value is calculated, and a curve is plotted between them. The area under this curve is an average precision of that iteration. At last, we will find the mean value of all obtained average precisions to get map value. We have observed that YOLO v3-Tiny is faster compared to YOLO v3, and therefore, it is more effective in real time target localization and identification of class. However, its accuracy decreases a bit in comparison to YOLO v3.



Fig.10: Experimental result-1 of YOLO v3



Fig.11: Experimental result-1 of YOLO v3-Tiny



Fig.12: Experimental result-2 of YOLO v3



Fig.13: Experimental result-2 of YOLO v3-Tiny

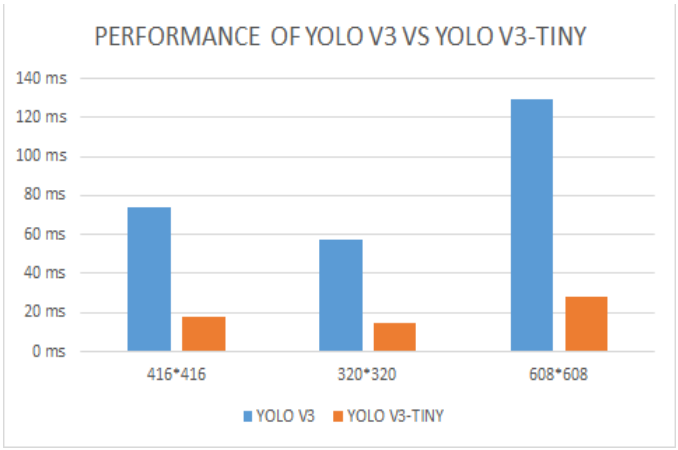


Fig.14: Speed Comparison

The above graph in Fig. 14 shows a comparison of the running time of YOLO v3 with YOLO v3-Tiny for different dimensions of images.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.

V. CONCLUSION

To identify and localize objects, there exist many methods with a trade-off in speed performance and accuracy of result. But yet we can't say any single algorithm is best over others. One can always select the method that suits the requirement at best. In a short period, object detection applications got much popularity and still a lot to cover in this area because of its vast scope of research[15] [18] [26]. This paper presents the comparison of various algorithms to identify and localize objects based on accuracy, time, and parameter values with varying sizes of the input image. We have identified a new methodology of single stage model for improving speed without sacrificing much accuracy. The comparison results show that YOLO v3-Tiny increases the speed of object detection while ensures the accuracy of the result. We can also extend object localization and recognition from static pictures to a video containing the dynamic sequence of images. We are planning to increase the accuracy of the localization of small targets in the future.

References

[1] A. Krizhevsky, I. Sutskever, and G. E.Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS*, 2012, doi: 10.1201/9781420010749.

[2] R. L. Galvez, A. A. Bandala, E. P. Dadios, R. R. P. Vicerra, and J. M. Z. Maningo, "Object Detection Using Convolutional Neural Networks," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-October, no. October, pp. 2023–2027, 2019, doi: 10.1109/TENCON.2018.8650517.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[4] R. Girshick, "Fast R-CNN," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 1440–1448, 2015, doi: 10.1109/ICCV.2015.169.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[6] P. Dong and W. Wang, "Better region proposals for pedestrian detection with R-CNN," 30th Anniv. Vis. Commun. Image Process., pp. 3–6, 2016, doi: 10.1109/VCIP.2016.7805452.

[7] W. Liu, D. Anguelov, D. Erhan, and C. Szegedy, "SSD: Single Shot MultiBox Detector," ECCV, vol. 1, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real- time object detection," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR, vol. 2017-Janua, pp. 6517–6525, 2017, doi: 10.1109/CVPR.2017.690.

[10] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv Prepr., 2018.

[11] S. Ding, F. Long, H. Fan, L. Liu, and Y. Wang, "A novel YOLOv3-tiny network for unmanned airship obstacle detection," IEEE 8th Data Driven Control Learn. Syst. Conf. DDCLS, pp. 277–281, 2019, doi: 10.1109/DDCLS.2019.8908875.

[12] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE CVPR, vol. 1, pp. 886–893, 2005, doi: 10.1109/CVPR.2005.177.

[13] C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, "Going Deeper with Convolutions," CVPR, 2015, doi: 10.1108/978-1-78973-723-320191012.

[14] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, 2013, doi: 10.1007/s11263-013-0620-5.

[15] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," IEEE Trans. Neural Networks Learn. Syst., vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," ECCV, pp. 346–361, 2014, doi: 10.1023/B:KICA.0000038074.96200.69.

[17] R. Nabati and H. Qi, "RRPN : RADAR REGION PROPOSAL NETWORK FOR OBJECT DETECTION IN AUTONOMOUS VEHICLES," IEEE Int. Conf. Image Process., pp. 3093–3097, 2019.

[18] L. Jiao et al., "A Survey of Deep Learning-Based Object Detection," IEEE Access, vol. 7, pp. 128837–128868, 2019, doi: 10.1109/access.2019.2939201.

[19] D. Wang, C. Li, S. Wen, X. Chang, S. Nepal, and Y. Xiang, "Daedalus: Breaking Non-Maximum Suppression in Object Detection via Adversarial Examples," arXiv Prepr., 2019.

[20] C. Ning, H. Zhou, Y. Song, and J. Tang, "Inception Single Shot MultiBox Detector for object detection," IEEE Int. Conf. Multimed. Expo Work. ICMEW, no. July, pp. 549–554, 2017, doi: 10.1109/ICMEW.2017.8026312.

[21] Z. Chen, R. Khemmar, B. Decoux, A. Atahouet, and J. Y. Ertaud, "Real time object detection, tracking, and distance and motion estimation based on deep learning: Application to smart mobility," 8th Int. Conf. Emerg. Secur. Technol. EST, pp. 1–6, 2019, doi: 10.1109/EST.2019.8806222.

[22] D. Xiao, F. Shan, Z. Li, B. T. Le, X. Liu, and X. Li, "A Target Detection Model Based on Improved Tiny-Yolov3 Under the Environment of Mining Truck," IEEE Access, vol. 7, pp. 123757–123764, 2019, doi: 10.1109/access.2019.2928603.

[23] Q. C. Mao, H. M. Sun, Y. B. Liu, and R. S. Jia, "Mini-YOLOv3: Real-Time Object Detector for Embedded Applications," IEEE Access, vol. 7, pp. 133529–133538, 2019, doi: 10.1109/ACCESS.2019.2941547.

[24] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A Real-time Object Detection Method for Constrained Environments," IEEE Access, vol. 8, pp. 1935–1944, 2019, doi10.1109/ACCESS.2019.2961959.

[25] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," *IEEE Int. Conf. Big Data, Big Data*, pp. 2503–2510, 2019, doi: 10.1109/BigData.2018.8621865.

[26] J. Huang *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," *30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, vol. 2017-Janua, pp. 3296–3305, 2017, doi: 10.1109/CVPR.2017.351.

[27] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, "Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3," *1st Int. Conf. Unmanned Veh. Syst.*, pp. 1–6, 2019, doi: 10.1109/UVS.2019.8658300.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 21,2020 at 08:35:23 UTC from IEEE Xplore. Restrictions apply.